The gamlss.family Distributions
Flexible Regression and Smoothing

Bob Rigby\(^1\)    Mikis Stasinopoulos\(^1\)

\(^1\)STORM, London Metropolitan University

XXV SIMPOSIO INTERNACIONAL DE ESTAD\'{E}STICA, Armenia, Colombia, August 2015
1 Types of distribution in GAMLSS

2 Continuous distributions
   - Summary of methods of generating distributions
   - Theoretical Comparison

3 Discrete Distributions
   - Count distributions
   - Examples
     - A stylometric application
     - The fish species data
   - Binomial response variables

4 Mixed Distributions

5 End
Information

Books & Articles

Three books on GAMLSS are in preparation:

1. Flexible Regression and Smoothing, the GAMLSS packages in R (first draft versions)
2. Distributions for Location Scale and Shape, the GAMLSS implementation in R (first draft versions under a old name)

The GAMLSS R reference card
Different types of distribution

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
Example of discrete distribution: Sichel

Sichel, SICHEL
SICHEL( mu = 5, sigma = 11.04, nu = 0.98 )

Sichel, SICHEL
SICHEL( mu = 5, sigma = 1.151, nu = 0 )

Sichel, SICHEL
SICHEL( mu = 5, sigma = 0.9602, nu = −1 )

Sichel, SICHEL
SICHEL( mu = 5, sigma = 43.9, nu = −3 )
Example of mixed distribution distributions: ZAGA

Zero adjusted GA

Zero adjusted Gamma c.d.f.
Types of continuous distributions

**negative skewness**

**positive skewness**

**platy–kurtosis**

**lepto–kurtosis**

**Figure**: Showing different types of continuous distributions
Two parameter continuous distributions in GAMLSS

Two parameter distributions

BE  Beta (0, 1)
GA  Gamma (0, ∞)
GU  Gumbel (−∞, ∞)
LO  Logistic (−∞, ∞)
LNO Log Normal (0, ∞)
NO  Normal (−∞, ∞)
IG  Inverse Gaussian (0, ∞)
RG  Reverse Gumbel (−∞, ∞)
WEI Weibull (also WEI2, WEI3) (0, ∞)
Three parameter continuous distributions in GAMLSS

Three parameter distributions

- **BCCG** Box-Cox Normal \((0, \infty)\)
- **exGAUS** Exponential-Gaussian \((-\infty, \infty)\)
- **GG** Generalized Gamma \((0, \infty)\)
- **GIG** Generalized Inverse Gaussian \((0, \infty)\)
- **PE** Power Exponential \((-\infty, \infty)\)
- **TF** \(t\) family \((-\infty, \infty)\)
- **SN1** skew normal type 1 \((-\infty, \infty)\)
- **SN2** skew normal type 2 \((-\infty, \infty)\)
Two and three parameters comparison: demos

1. demo.PE.NO()
2. demo.TF.NO()
Four parameter continuous distributions in GAMLSS

Four parameter distributions

- **BCT**  Box-Cox \(t\) \((0, \infty)\)
- **BCPE**  Box-Cox Power Exponential \((0, \infty)\)
- **EGB2**  Exponential Generalized Beta type 2 \((-\infty, \infty)\)
- **GT**  Generalized \(t\) \((-\infty, \infty)\)
- **JSU**  Johnson Su \((-\infty, \infty)\)
- **SHASH**  Sinh Arc-Sinh \((-\infty, \infty)\)
- **SEP1-SEP4**  Skew Exponential Power \((-\infty, \infty)\)
- **ST1-ST5**  Skew \(t\) \((-\infty, \infty)\)
- **SST**  Shew \(t\) \((-\infty, \infty)\) reparation of ST3
## Continuous 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>
Continuous GAMLSS family distributions defined on \((-\infty, +\infty)\)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Formulas</th>
<th>Parameter(s)</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>
The NET distribution

The NET distribution is a continuous probability distribution. It is a mixture of a normal distribution and a t-distribution, with the normal distribution representing the central part of the distribution and the t-distribution representing the tails. The NET distribution is defined by two parameters: the degrees of freedom for the t-distribution and the location parameter.

The NET distribution is useful in situations where the data is expected to have heavier tails than a normal distribution, but not as heavy as a t-distribution. It can be used in various fields such as finance, economics, and engineering.
Symmetric SEP1 distribution \( (\nu = 0) \)
Skew SEP1 distributions ($\tau = 0.5$)
Skew SEP1 distributions ($\tau = 2$)
Skew SEP1 distributions ($\tau = 10$)
Skew SEP1 distributions ($\tau = 1000$)
## 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>
Positive response: 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

\[ d \log T(x, \mu = 0) \]

\[ p \log T(x, \mu = 0) \]

Histogram of Y

\[ q \log T(x, \mu = 0) \]
Positive response: demos

1. demo.NO.LOGNO()
2. demo.BCPE()
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>
Continuous GAMLSS family distributions defined on $(0, 1)$: 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 GAMLSS family distributions defined on (0, 1): logit T distribution
Four parameters from 0 to 1: demos

demo.GB1()
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
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 $t$</td>
<td>SB($\mu, \sigma, \nu, \tau$)</td>
</tr>
<tr>
<td></td>
<td></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>JSU0($\mu, \sigma, \nu, \tau$)</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>BCCG($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Box-Cox power exponential</td>
<td>BCPE($\mu$, $\sigma$, $\nu$, $\tau$)</td>
</tr>
<tr>
<td></td>
<td>Box-Cox t</td>
<td>BCT($\mu$, $\sigma$, $\nu$, $\tau$)</td>
</tr>
<tr>
<td></td>
<td>Generalized beta type 2</td>
<td>GB2($\mu$, $\sigma$, $\nu$, $\tau$)</td>
</tr>
<tr>
<td></td>
<td>Generalized gamma</td>
<td>GG($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Inverse gamma</td>
<td>IGA($\mu$, $\sigma$)</td>
</tr>
<tr>
<td></td>
<td>log $t$</td>
<td>LOGT($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Pareto Type 2</td>
<td>PA2O($\mu$, $\sigma$)</td>
</tr>
<tr>
<td>$k_1 = 2$</td>
<td>Box-Cox Cole-Green</td>
<td>BCCG($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>LOGNO($\mu$, $\sigma$)</td>
</tr>
<tr>
<td></td>
<td>Log Weibull</td>
<td>LOGWEI($\mu$, $\sigma$)</td>
</tr>
<tr>
<td>$1 \leq k_1 &lt; \infty$</td>
<td>Box-Cox power exponential</td>
<td>BCPE($\mu$, $\sigma$, $\nu$, $\tau$)</td>
</tr>
<tr>
<td>$0 &lt; k_3 &lt; \infty$</td>
<td>Box-Cox Cole-Green</td>
<td>BCCG($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Box-Cox power exponential</td>
<td>BCPE($\mu$, $\sigma$, $\nu$, $\tau$)</td>
</tr>
<tr>
<td></td>
<td>Generalized gamma</td>
<td>GG($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Weibull</td>
<td>WEI($\mu$, $\sigma$)</td>
</tr>
<tr>
<td>$k_3 = 1$</td>
<td>Exponential</td>
<td>EX($\mu$)</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>GA($\mu$, $\sigma$)</td>
</tr>
<tr>
<td></td>
<td>Generalized inverse Gaussian</td>
<td>GIG($\mu$, $\sigma$, $\nu$)</td>
</tr>
<tr>
<td></td>
<td>Inverse Gaussian</td>
<td>IG($\mu$, $\sigma$)</td>
</tr>
</tbody>
</table>
How to fit distributions in R

`optim()` or `mle()` requires initial parameter values

`gamlssML()` mle a using variation of the `mle()` function

`gamlss()` mle using RS or CG or mixed algorithms,

`histdist()` 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:

- **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

glass strength

Frequency

0.5 1.0 1.5 2.0
0 5 10 15 20
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
Strength of glass fibres data: fitting the models

```r
> m1<-fitDist(strength, data=glass, k=2, extra="TFtr")
  # AIC
> m2<-fitDist(strength, data=glass, k=log(length(strength)),
  extra="TFtr")  # SBC
> m1$fit[1:8]
  SEP4      SEP3    SHASHo     EGB2      JSU      BCPEo
> m2$fit[1:8]
  SEP4      SEP3    SHASHo     EGB2      GU      WEI3
36.26044  36.55444  36.59054  36.95945  38.19839  38.69995  38.9842
```
Strength of glass fibres data: Results

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

SEP4 distribution
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
Global Deviance: 19.8111
AIC: 27.8111
Count distributions

The three major problems encountered when modelling count data using the Poisson distribution.

- overdispersion
- excess (or shortage) of zero values
- long tails (rare events)
## Discrete distribution modelling

<table>
<thead>
<tr>
<th>Par.</th>
<th>Modelling</th>
<th>Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Location</td>
<td>PO</td>
</tr>
<tr>
<td>2</td>
<td>Location and scale</td>
<td>NBI, NBII, PIG</td>
</tr>
<tr>
<td>2</td>
<td>Location and zero probability</td>
<td>ZALG, ZAP, ZIP, ZIP2</td>
</tr>
<tr>
<td>3</td>
<td>Location, scale and skewness</td>
<td>DEL, SI, SICHEL</td>
</tr>
<tr>
<td>3</td>
<td>Location, scale and zero probability</td>
<td>ZANBI, ZINBI, ZIPIG</td>
</tr>
</tbody>
</table>
Different count data distributions

- ZIP(10,0.5)
- NB(5,1)
- NBtr(2.77,2.61)
- PIG(5,1)
Zero inflated distributions

Zero inflated distribution, $Y \sim \text{ZID}$ is given by

$Y = 0$ with probability $p$

$Y \sim D$ with probability $1 - p$.

Hence

$$P(Y = y) = \begin{cases} 
  p + (1 - p)P(Y_1 = 0) & \text{if } y = 0 \\
  (1 - p)P(Y_1 = y) & \text{if } y = 1, 2, 3, \ldots
\end{cases}$$

where $Y_1 \sim D$. 

49
ZINBI distribution plots

Zero inflated negative binomial type I, ZINBI
ZINBI( mu = 9, sigma = 0.1111, nu = 0.4444 )

Zero inflated negative binomial type I, ZINBI
ZINBI( mu = 7, sigma = 0.4286, nu = 0.2857 )

Zero inflated negative binomial type I, ZINBI
ZINBI( mu = 6, sigma = 0.6667, nu = 0.1667 )

Zero inflated negative binomial type I, ZINBI
ZINBI( mu = 5.5, sigma = 0.8182, nu = 0.09091 )
Zero adjusted distributions

Zero adjusted distribution, \( Y \sim ZAD \) is given by
\( Y = 0 \) with probability \( p \)
\( Y \sim Dtr \) with probability \( 1 - p \),
where \( Dtr \) is a truncated distribution, \( D \) truncated at zero.
Hence

\[
P(Y = y) = \begin{cases} 
    p & \text{if } y = 0 \\
    (1 - p) \frac{P(Y_1 = y)}{1 - P(Y_1 = 0)} & \text{if } y = 1, 2, 3, \ldots 
\end{cases}
\]

(1)

where \( Y_1 \sim D \).
ZANBI distribution plots

Zero altered negative binomial type I, ZANBI
ZANBI( μ = 5, σ = 1, ν = 0.1667 )

Zero altered negative binomial type I, ZANBI
ZANBI( μ = 4, σ = 1.5, ν = 0.09159 )

Zero altered negative binomial type I, ZANBI
ZANBI( μ = 3, σ = 2.33, ν = 0.01695 )

Zero altered negative binomial type I, ZANBI
ZANBI( μ = 2.8, σ = 2.57, ν = 0.00213 )
Different (overdispersed) count data approaches

(a) *Ad-hoc* solutions

(i) quasi-likelihood (QL), Extended QL
(ii) Efron’s Double Exponential
(iii) pseudo-likelihood (PL)

(b) Discretized continuous distributions
For example if $F_W(w)$ is the cdf a continuous random variable $W$ defined in $\mathbb{R}^+$ then $f_Y(y) = F_W(y + 1) - F_W(y)$

(c) Random effect at the observation level solutions.
$f_Y(y) = \int f(y|\gamma)f_\gamma(\gamma)d\gamma$. 
(c) Random effect at the observation level

(i) when an explicit continuous mixture distribution, \( f_Y(y) \), exists.

(ii) when a continuous mixture distribution, \( f_Y(y) \), is not explicit but is approximated by integrating out the random effect using approximations, e.g. Gaussian quadrature or Laplace approximation.

(iii) when a 'non-parametric' mixture (effectively a finite mixture) is assumed for the response variable.
Random effect at the observation level case (i)

(i) Explicit continuous mixture distribution

\[
f_Y(y) = \int f(y|\gamma) f_\gamma(\gamma) \, d\gamma
\]

- \( Y \sim NBI(\mu, \sigma) \)
- \( Y|\gamma \sim PO(\gamma\mu) \)
- \( \gamma \sim GA(1, \sigma^{1/2}) \)
Random effect at the observation level case (ii)

(ii) Non-explicit continuous mixture distribution

\[ f_Y(y) = \int f(y|\gamma) f(\gamma) \, d\gamma \]

- \( Y \sim PO - Normal(\mu, \sigma) \)
- \( Y|\gamma \sim PO(\gamma \mu) \)
- \( \log(\gamma) \sim NO(1, \sigma) \)
Random effect at the observation level case (iii)

(iii) Non-parametric mixture distribution

\[
\begin{align*}
   f_Y(y) &= \sum_{k=1}^{K} f(y|\gamma_k) p(\gamma = \gamma_k) \\
   \text{discrete} & \quad \text{discrete} & \quad \text{continuous}
\end{align*}
\]

- \( Y \sim PO - NPFM(\mu, \sigma) \)
- \( Y|\gamma \sim PO(\gamma\mu) \)
- \( \log(\gamma) \sim NPFM(2) \)

where NPFM(2) equals Non-Parametric Finite Mixture with 2 point probabilities
## Explicit continuous mixture distribution

<table>
<thead>
<tr>
<th>Distributions</th>
<th>R Name</th>
<th>mixing distribution for $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>PO$(\mu)$</td>
<td>-</td>
</tr>
<tr>
<td>Neg. bin. I</td>
<td>NBI$(\mu, \sigma)$</td>
<td>GA$(1, \sigma^{\frac{1}{2}})$</td>
</tr>
<tr>
<td>Neg. bin. II</td>
<td>NBII$(\mu, \sigma)$</td>
<td>GA$(1, \sigma^{\frac{1}{2}}/\mu)$</td>
</tr>
<tr>
<td>Poisson IG</td>
<td>PIG$(\mu, \sigma)$</td>
<td>IG$(1, \sigma^{\frac{1}{2}})$</td>
</tr>
<tr>
<td>Sichel</td>
<td>SICHEL$(\mu, \sigma, \nu)$</td>
<td>GIG$(1, \sigma^{\frac{1}{2}}, \nu)$</td>
</tr>
<tr>
<td>Delaporte</td>
<td>DEL$(\mu, \sigma, \nu)$</td>
<td>SG$(1, \sigma^{\frac{1}{2}}, \nu)$</td>
</tr>
<tr>
<td>Zero inflated Poisson</td>
<td>ZIP$(\mu, \sigma)$</td>
<td>BI$(1, 1 - \sigma)$</td>
</tr>
<tr>
<td>Zero inflated Poisson 2</td>
<td>ZIP2$(\mu, \sigma)$</td>
<td>$(1 - \sigma)^{-1}$BI$(1, 1 - \sigma)$</td>
</tr>
<tr>
<td>Zero inflated neg. bin.</td>
<td>ZINBI$(\mu, \sigma, \nu)$</td>
<td>zero inflated gamma</td>
</tr>
<tr>
<td>Poisson-Tweedie</td>
<td>-</td>
<td>Tweedie family</td>
</tr>
</tbody>
</table>
Table: Discrete gamlss family distributions for count data

<table>
<thead>
<tr>
<th>R Name</th>
<th>params</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO(µ)</td>
<td>1</td>
<td>µ</td>
<td>µ</td>
</tr>
<tr>
<td>NBI(µ, σ)</td>
<td>2</td>
<td>µ</td>
<td>µ + σµ²</td>
</tr>
<tr>
<td>NBII(µ, σ)</td>
<td>2</td>
<td>µ</td>
<td>µ + σµ</td>
</tr>
<tr>
<td>PIG(µ, σ)</td>
<td>2</td>
<td>µ</td>
<td>µ + σµ²</td>
</tr>
<tr>
<td>SICHEL(µ, σ, ν)</td>
<td>3</td>
<td>µ</td>
<td>µ + h(σ, ν)µ²</td>
</tr>
<tr>
<td>DEL(µ, σ, ν)</td>
<td>3</td>
<td>µ</td>
<td>µ + σ(1 - ν)²µ²</td>
</tr>
<tr>
<td>ZIP(µ, σ)</td>
<td>2</td>
<td>(1 - σ)µ</td>
<td>(1 - σ)µ + σ(1 - σ)µ²</td>
</tr>
<tr>
<td>ZIP2(µ, σ)</td>
<td>2</td>
<td>µ</td>
<td>µ + σ(1 - σ)µ²</td>
</tr>
</tbody>
</table>
Comparison of the marginal distributions using a (ratio moment) diagram of their skewness and kurtosis.
A stylometric application

Data summary:

R data file: stylo in package `gamlss.data` of dimensions 64 \( \times \) 2

source: Dr Mario Corina-Borja

variables

word : is the number of times a word appears in a single text

freq : the frequency of the number of times a word appears in a text

purpose: to demonstrate the fitting of a truncated discrete dist.

conclusion: the truncated SICHEL distributions fits best
A stylometric application

```r
library(gamlss.tr)
data(stylo)
plot(freq ~ word, data = stylo, type = "h", xlim = c(0, 22), xlab = "no of times", ylab = "frequencies", col = "blue")
```
The stylometric data
A stylometric application

> library(gamlss.tr)
> gen.trun(par = 0, family = PO, type = "left")
A truncated family of distributions from PO has been generated and saved under the names:
dPOtr pP0tr qPOtr rP0tr P0tr
The type of truncation is left and the truncation parameter is 0.
> gen.trun(par = 0, family = NBII, type = "left")
...
> gen.trun(par = 0, family = DEL, type = "left")
...
> gen.trun(par = 0, family = SICHEL, type = "left",
  +  delta = 0.001)
...
A stylometric application

> mPO <- gamlss(word ~ 1, weights = freq, data = stylo, + family = P0tr, trace = FALSE)
> mNBII <- gamlss(word ~ 1, weights = freq, data = stylo, + family = NBIIt, n.cyc = 50, trace = FALSE)
> mDEL <- gamlss(word ~ 1, weights = freq, data = stylo, + family = DELt, n.cyc = 50, trace = FALSE)
> mSI <- gamlss(word ~ 1, weights = freq, data = stylo, + family = SICHELt, n.cyc = 50, trace = FALSE)
> GAIC(mPO, mNBII, mDEL, mSI)

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mSI</td>
<td>3</td>
<td>5148.454</td>
</tr>
<tr>
<td>mDEL</td>
<td>3</td>
<td>5160.581</td>
</tr>
<tr>
<td>mNBII</td>
<td>2</td>
<td>5311.627</td>
</tr>
<tr>
<td>mPO</td>
<td>1</td>
<td>9207.459</td>
</tr>
</tbody>
</table>
The stylometric data

(b) Poisson

(c) negative binomial II

(c) Delaporte

(d) Sichel
The fish species data

**Data summary:** the fish species data

**R data file:** species in package gamlss.data of dimensions $70 \times 2$

**variables**

- **fish**: the number of different species in 70 lakes in the world
- **lake**: the lake area
The fish species data

![Graph showing the relationship between log(lake) and fish species count.](image)
The fish species data

There are several questions that need to be answered.

- How does the mean of $y$ depend on $x$?
- Is $y$ overdispersed Poisson?
- How does the variance of $y$ depend on its mean?
- What is the distribution of $y$ given $x$?
- Do the scale and shape parameters of the distribution of $y$ depend on $x$?
Overdispersed count data approaches

<table>
<thead>
<tr>
<th>Model</th>
<th>$f_Y(y)$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\nu$</th>
<th>$DEV$</th>
<th>$df$</th>
<th>$AIC$</th>
<th>$SBC$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PO</td>
<td>$x &lt; 2 &gt;$</td>
<td>-</td>
<td>-</td>
<td>1849.3</td>
<td>3</td>
<td>1855.3</td>
<td>1862.0</td>
</tr>
<tr>
<td>2</td>
<td>NBI</td>
<td>$x$</td>
<td>1</td>
<td>-</td>
<td>619.8</td>
<td>3</td>
<td>625.8</td>
<td>632.6</td>
</tr>
<tr>
<td>3</td>
<td>NBI</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>-</td>
<td>614.3</td>
<td>4</td>
<td>622.3</td>
<td>631.3</td>
</tr>
<tr>
<td>4</td>
<td>NBI</td>
<td>$cs(x, 3)$</td>
<td>1</td>
<td>-</td>
<td>611.9</td>
<td>6</td>
<td>623.9</td>
<td>637.4</td>
</tr>
<tr>
<td>5</td>
<td>NBI</td>
<td>$x &lt; 2 &gt;$</td>
<td>$x$</td>
<td>-</td>
<td>605.0</td>
<td>5</td>
<td>615.0</td>
<td>626.2</td>
</tr>
<tr>
<td>6</td>
<td>NBI-fam</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>1</td>
<td>606.0</td>
<td>5</td>
<td>616.0</td>
<td>627.3</td>
</tr>
<tr>
<td>7</td>
<td>NBI-fam</td>
<td>$x &lt; 2 &gt;$</td>
<td>$x$</td>
<td>1</td>
<td>604.9</td>
<td>6</td>
<td>616.9</td>
<td>630.4</td>
</tr>
</tbody>
</table>

Table: Comparison of models for the fish species data
## Overdispersed count data approaches

<table>
<thead>
<tr>
<th>Model</th>
<th>$f_Y(y)$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\nu$</th>
<th>DEV</th>
<th>df</th>
<th>AIC</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>PIG</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>-</td>
<td>613.3</td>
<td>4</td>
<td>621.3</td>
<td>630.3</td>
</tr>
<tr>
<td>9</td>
<td>SI</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>$x$</td>
<td>597.7</td>
<td>6</td>
<td>609.7</td>
<td>623.2</td>
</tr>
<tr>
<td>10</td>
<td>DEL</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>$x$</td>
<td>600.6</td>
<td>6</td>
<td>612.6</td>
<td>626.1</td>
</tr>
<tr>
<td>11</td>
<td>DEL</td>
<td>$x &lt; 2 &gt;$</td>
<td>-</td>
<td>$x$</td>
<td>600.6</td>
<td>5</td>
<td>610.6</td>
<td>621.9</td>
</tr>
<tr>
<td>12</td>
<td>PO-Normal</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>-</td>
<td>615.2</td>
<td>4</td>
<td>623.2</td>
<td>632.2</td>
</tr>
<tr>
<td>13</td>
<td>NBI-Normal</td>
<td>$x &lt; 2 &gt;$</td>
<td>$x$</td>
<td>1</td>
<td>603.7</td>
<td>6</td>
<td>615.7</td>
<td>629.2</td>
</tr>
<tr>
<td>14</td>
<td>PO-NPFM(5)</td>
<td>$x &lt; 2 &gt;$</td>
<td>-</td>
<td>-</td>
<td>601.9</td>
<td>13</td>
<td>627.9</td>
<td>657.2</td>
</tr>
<tr>
<td>15</td>
<td>NB-NPFM(2)</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>-</td>
<td>611.9</td>
<td>6</td>
<td>623.9</td>
<td>637.4</td>
</tr>
<tr>
<td>16</td>
<td>doublePO</td>
<td>$x &lt; 2 &gt;$</td>
<td>$x$</td>
<td>-</td>
<td>616.4</td>
<td>5</td>
<td>626.4</td>
<td>637.6</td>
</tr>
<tr>
<td>17</td>
<td>IGdisc</td>
<td>$x &lt; 2 &gt;$</td>
<td>1</td>
<td>-</td>
<td>603.3</td>
<td>4</td>
<td>611.3</td>
<td>620.3</td>
</tr>
</tbody>
</table>
Fitted mean of the Sichel distribution
Fitted Sichel distributions for observations (a) 40 and (b) 67

Sichel, SICHEL

SICHEL( mu = 22.61, sigma = 1.447, nu = -5.646 )

Sichel, SICHEL

SICHEL( mu = 47.67, sigma = 1.447, nu = -1.211 )
There are only two distributions here
- binomial
- beta binomial
Mixed distributions

A mixed distribution is a mixture of two components:

- a continuous distribution and
- a discrete distribution

i.e. it is a continuous distribution where the range of $Y$ also includes discrete values with non-zero probabilities.
Zero adjusted distributions on zero and the positive real line \([0, \infty)\)

They are a mixture of a discrete value 0 with probability \(p\), and a continuous distribution on the positive real line \((0, \infty)\) with probability \((1 - p)\).

The probability (density) function of \(Y\) is \(f_Y(y)\) given by

\[
f_Y(y) = \begin{cases} 
    p & \text{if } y = 0 \\
    (1 - p)f_W(y) & \text{if } y > 0
\end{cases}
\]  

(2)

for \(0 \leq y < \infty\), where \(0 < p < 1\) and \(f_W(y)\) is a probability density function defined on \((0, \infty)\), i.e. for \(0 < y < \infty\).
Zero adjusted gamma distribution

Zero adjusted GA

Zero adjusted Gamma c.d.f.
Zero adjusted gamma distribution, \textbf{ZAGA}(\mu, \sigma, \nu)

The default link functions relating the parameters \((\mu, \sigma, \nu)\) to the predictors \((\eta_1, \eta_2, \eta_3)\), which may depend on explanatory variables, are

\[
\begin{align*}
\log \mu &= \eta_1 \\
\log \sigma &= \eta_2 \\
\log \left( \frac{\nu}{1 - \nu} \right) &= \eta_3.
\end{align*}
\]
Zero adjusted gamma distribution, \textbf{ZAGA}(\mu, \sigma, \nu)

The ZAGA model is equivalent to

- a gamma distribution \textbf{GA}(\mu, \sigma) model for \( Y > 0 \) together
- with a binary model for recoded variable \( Y_1 \) given by

\[
Y_1 = \begin{cases} 
0 & \text{if } Y > 0 \\
1 & \text{if } Y = 0
\end{cases}
\] (3)

i.e.

\[
p(Y_1 = y_1) = \begin{cases} 
(1 - \nu) & \text{if } y_1 = 0 \\
\nu & \text{if } y_1 = 1
\end{cases}
\] (4)
Distributions on the interval (0,1) inflated at 0 and 1

These distributions are appropriate when the response variable $Y$ takes values from 0 to 1 including 0 and 1, i.e. range $[0,1]$. They are a mixture of three components:

- a discrete value 0 with probability $p_0$,
- a discrete value 1 with probability $p_1$,
- and a continuous distribution on the unit interval $(0, 1)$ with probability $(1 - p_0 - p_1)$. 
Distributions on the interval \((0,1)\) inflated at 0 and 1

The probability (density) function of \(Y\) is \(f_Y(y)\) given by

\[
f_Y(y) = \begin{cases} 
  p_0 & \text{if } y = 0 \\
  (1 - p_0 - p_1)f_W(y) & \text{if } 0 < y < 1 \\
  p_1 & \text{if } y = 1 
\end{cases} \tag{5}
\]

or \(0 \leq y \leq 1\), where \(0 < p_0 < 1\), \(0 < p_1 < 1\) and \(0 < p_0 + p_1 < 1\) and \(f_W(y)\) is a probability density function defined on \((0,1)\), i.e. for \(0 < y < 1\).
Beta Inflated distribution

**Probability density function**

**Cumulative distribution function**
END

for more information see

www.gamlss.org