The gamlss.family Distributions

Flexible Regression and Smoothing

Bob Rigby[1] Mikis Stasinopoulos[1]

[1]STORM, London Metropolitan University

XXV SIMPOSIO INTERNACIONAL DE ESTADêSTICA, Armenia, Colombia, August 2015

Information

gamlss

HOME MORE ON GAMLSS THE TEAM

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)
3. Generalised Additive Models for Location Scale and Shape: A Distributional Regression Approach.

The GAMLSS R reference card

Different types of distribution

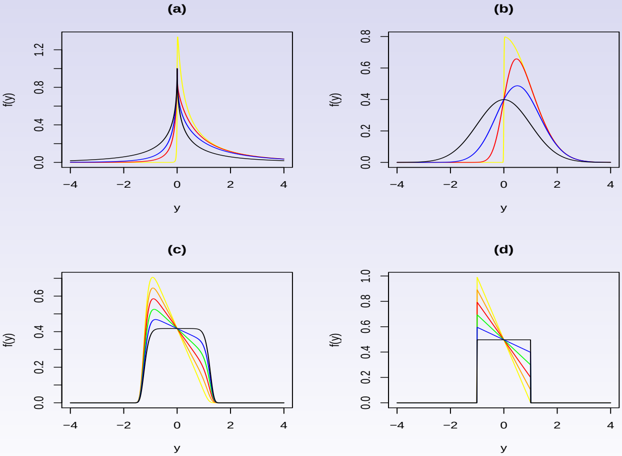
1. *continuous distributions*: $f_Y(y|\boldsymbol{\theta})$, are usually defined on $(-\infty, +\infty)$, $(0, +\infty)$ or $(0, 1)$.
2. *discrete distributions*: $P(Y = y|\boldsymbol{\theta})$ are defined on $y = 0, 1, 2, \ldots, n$, where n is a known finite value or n is infinite, i.e. usually discrete (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.

gamlss

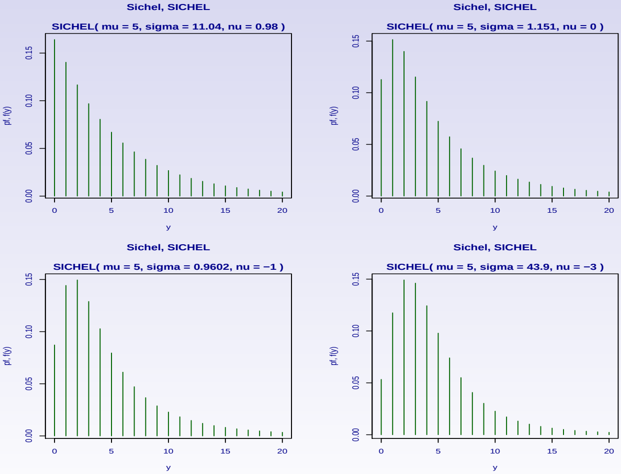
Different types of distribution: demos

1. demo.GA()
2. demo.PO()
3. demo.BI()
4. demo.BE()
5. demo.BEINF()

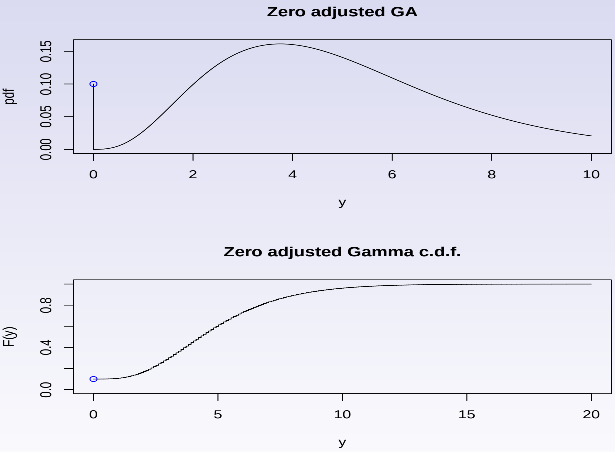
Example of continuous distribution



Example of discrete distribution: Sichel



Example of mixed distribution distributions: ZAGA



gamlss

Types of continuous distributions

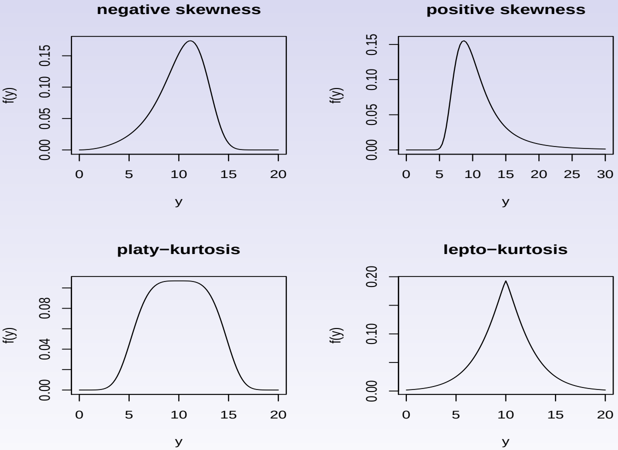


Figure : Showing different types of continuous distributions

Two parameter continuous distributions in GAMLSS

Two parameter distributions

BE	Beta $(0, 1)$
GA	Gamma $(0, \infty)$
GU	Gumbel $(-\infty, \infty)$
LO	Logistic $(-\infty, \infty)$
LNO	Log Normal $(0, \infty)$
NO	Normal $(-\infty, \infty)$
IG	Inverse Gaussian $(0, \infty)$
RG	Reverse Gumbel $(-\infty, \infty)$
WEI	Weibull (also WEI2, WEI3) $(0, \infty)$

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 $(-\infty, \infty)$ repar. of ST3

gamlss

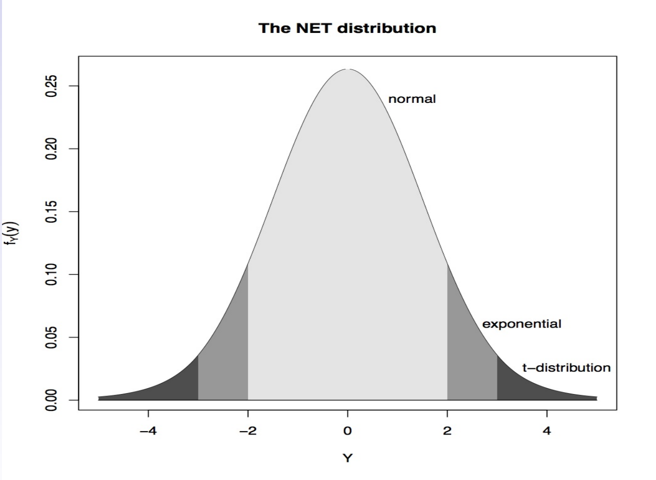
Continuous GAMLSS family distributions defined on $(-\infty, +\infty)$

Distributions	family	no par.	skewness	kurtosis
Exp. Gaussian	exGAUS	3	positive	-
Exp.G. beta 2	EGB2	4	both	lepto
Gen. t	GT	4	(symmetric)	lepto
Gumbel	GU	2	(negative)	-
Johnson's SU	JSU, JSUo	4	both	lepto
Logistic	LO	2	(symmetric)	(lepto)
Normal-Expon.-t	NET	2,(2)	(symmetric)	lepto
Normal	NO-NO2	2	(symmetric)	
Normal Family	NOF	3	(symmetric)	(meso)

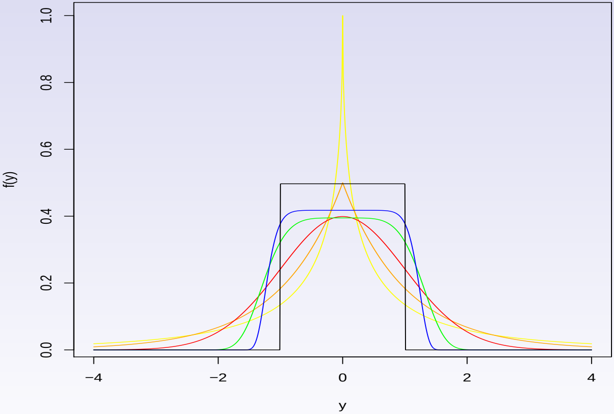
Continuous GAMLSS family distributions defined on $(-\infty, +\infty)$

Power Expon.	PE-PE2	3	(symmetric)	both
Reverse Gumbel	RG	2	positive	
Sinh Arcsinh	SHASH,	4	both	both
Skew Exp. Power	SEP1-SEP4	4	both	both
Skew *t*	ST1-ST5, SST	4	both	lepto
t Family	TF	3	(symmetric)	lepto

The NET distribution

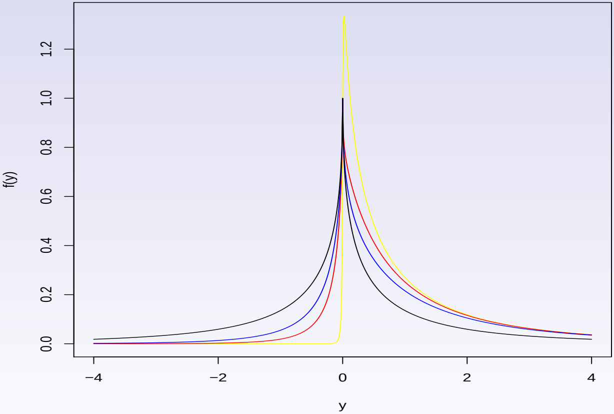


Symmetric SEP1 distribution ($\nu = 0$)



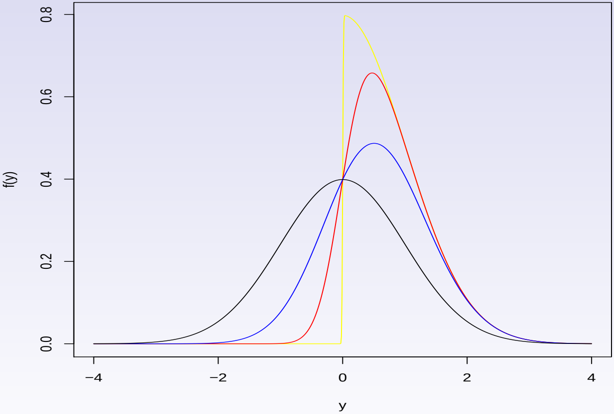
gamlss

Skew SEP1 distributions ($\tau = 0.5$)

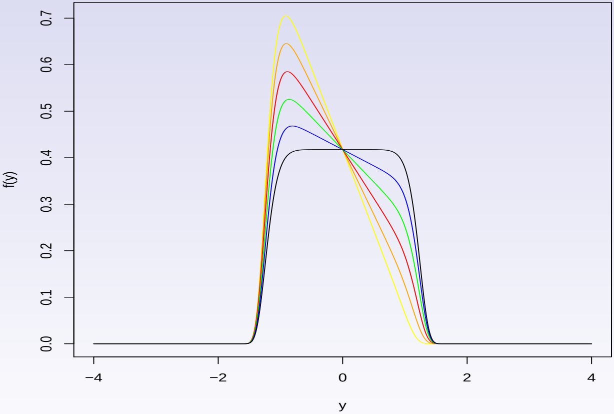


gamlss

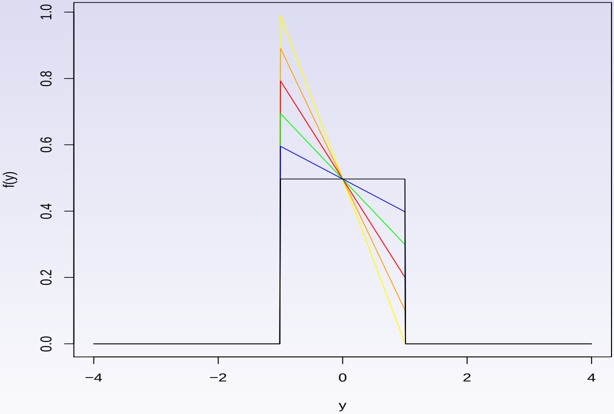
Skew SEP1 distributions ($\tau = 2$)



Skew SEP1 distributions ($\tau = 10$)



Skew SEP1 distributions ($\tau = 1000$)



Continuous GAMLSS family distributions defined on $(0, +\infty)$

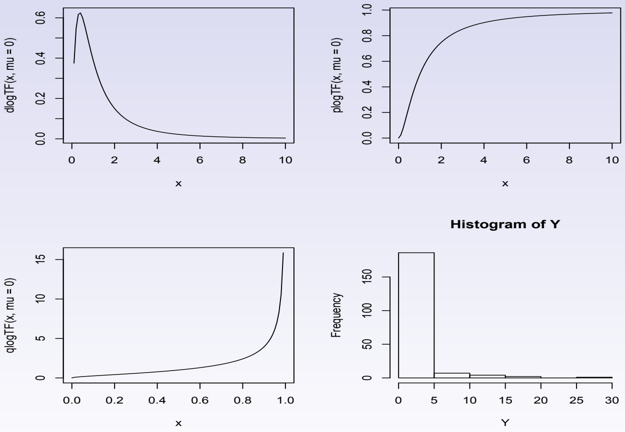
Distributions	family	no par.	skewness	kurtosis
BCCG	BCCG	3	both	
BCPE	BCPE	4	both	both
BCT	BCT	4	both	lepto
Exponential	EXP	1	(positive)	-
Gamma	GA	2	(positive)	-
Gen. Beta type 2	GB2	4	both	both
Gen. Gamma	GG-GG2	3	positive	-
Gen. Inv. Gaussian	GIG	3	positive	-
Inv. Gaussian	IG	2	(positive)	-
Log Normal	LOGNO	2	(positive)	-
Log Normal family	LNO	2,(1)	positive	
Reverse Gen. Extreme	RGE	3	positive	-
Weibull	WEI-WEI3	2	(positive)	-

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")`

gamlss

Positive response: log T distribution



gamlss

Positive response: demos

1. demo.NO.LOGNO()
2. demo.BCPE()

Continuous GAMLSS family distributions defined on $(0, 1)$

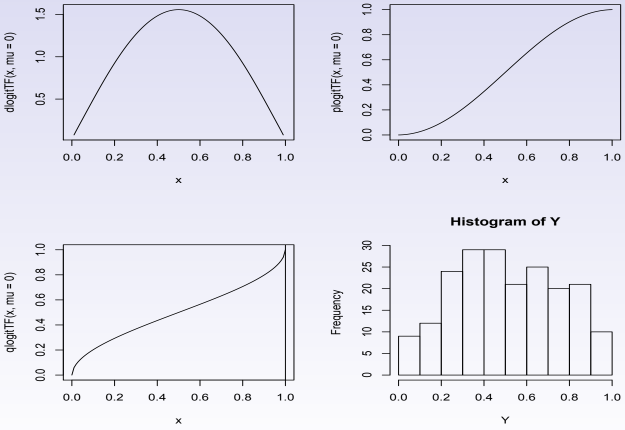
Distributions	family	no par.	skewness	kurtosis
Beta	BE	2	(both)	-
Beta original	BEo	2	(both)	-
Logit normal	LOGITNO	2	(both)	-
Generalized beta type 1	GB1	4	(both)	(both)

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")`

gamlss

Continuous GAMLSS family distributions defined on $(0, 1)$: logit T distribution



Four parameters from 0 to 1: demos

1. 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 μ, σ, ν and τ
7. Flexibility in modelling skewness and kurtosis
8. Range of Y

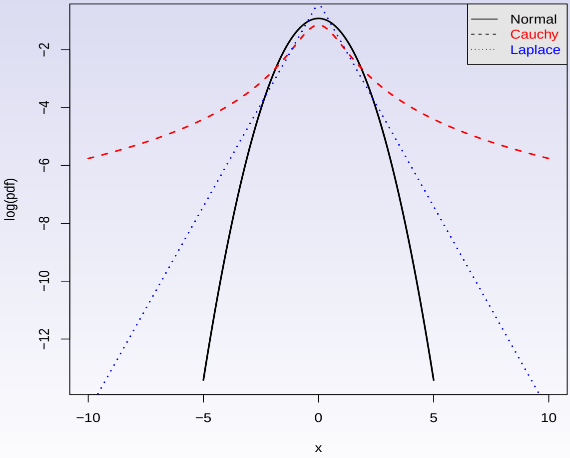
Theoretical comparison of distributions

- 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?

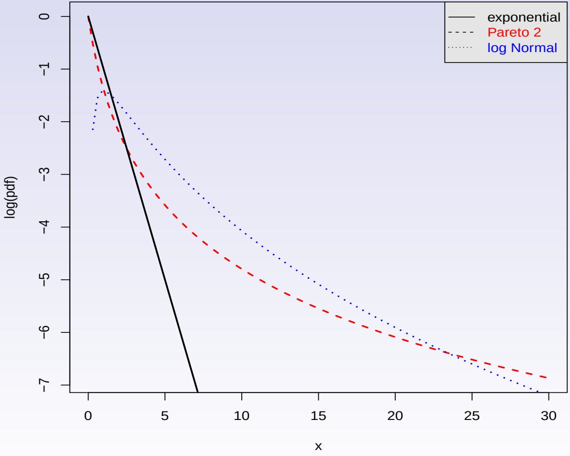
- We can compare the tail behaviour in of the distributions
- Since most of then are location-scale we can compare their flexibility in terms of skewness and kurtosis

Comparing tails: real line



gamlss

Comparing tails: positive real line



gamlss

Types of tails

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

$$
\begin{aligned}
\text{Type I:} &\quad -k_2 \, (\log |y|)^{k_1}, \\
\text{Type II:} &\quad -k_4 \, |y|^{k_3}, \\
\text{Type III:} &\quad -k_6 \, e^{k_5 |y|},
\end{aligned}
$$

decreasing k's results in a heavier tail,

Types of tails: classification (real line)

Value of k_1–k_6	Distribution name	Distribution
$k_1 = 1$	Cauchy	$\text{CA}(\mu, \sigma)$
	Generalized t	$\text{GT}(\mu, \sigma, \nu, \tau)$
	Skew t type 3	$\text{ST3}(\mu, \sigma, \nu, \tau)$
	Skew t type 4	$\text{ST4}(\mu, \sigma, \nu, \tau)$
	Stable	$\text{SB}(\mu, \sigma, \nu, \tau)$
	t	$\text{TF}(\mu, \sigma, \nu)$
$k_1 = 2$	Johnson's SU	$\text{JSU}(\mu, \sigma, \nu, \tau)$
	Johnson's SU original	$\text{JSUo}(\mu, \sigma, \nu, \tau)$
$0 < k_3 < \infty$	Power exponential	$\text{PE}(\mu, \sigma, \nu)$
	Power exponential type 2	$\text{PE2}(\mu, \sigma, \nu)$
	Sinh-arcsinh original	$\text{SHASHo}(\mu, \sigma, \nu, \tau)$
	Sinh-arcsinh	$\text{SHASH}(\mu, \sigma, \nu, \tau)$
	Skew exponential power type 3	$\text{SEP3}(\mu, \sigma, \nu, \tau)$
	Skew exponential power type 4	$\text{SEP4}(\mu, \sigma, \nu, \tau)$
$k_3 = 1$	Exponential generalized beta type 2	$\text{EGB2}(\mu, \sigma, \nu, \tau)$
	Gumbel	$\text{GU}(\mu, \sigma)$
	Laplace	$\text{LA}(\mu, \sigma)$
	Logistic	$\text{LG}(\mu, \sigma)$
	Reverse Gumbel	$\text{RG}(\mu, \sigma)$
$k_3 = 2$	Normal	$\text{NO}(\mu, \sigma)$
$0 < k_5 < \infty$	Gumbel	$\text{GU}(\mu, \sigma)$
	Reverse Gumbel	$\text{RG}(\mu, \sigma)$

Types of tails: classification (positive real line)

Value of k_1–k_6	Distribution name	Distribution
$k_1 = 1$	Box-Cox Cole-Green	$\text{BCCG}(\mu, \sigma, \nu)$
	Box-Cox power exponential	$\text{BCPE}(\mu, \sigma, \nu, \tau)$
	Box-Cox t	$\text{BCT}(\mu, \sigma, \nu, \tau)$
	Generalized beta type 2	$\text{GB2}(\mu, \sigma, \nu, \tau)$
	Generalized gamma	$\text{GG}(\mu, \sigma, \nu)$
	Inverse gamma	$\text{IGA}(\mu, \sigma)$
	log t	$\text{LOGT}(\mu, \sigma, \nu)$
	Pareto Type 2	$\text{PA2o}(\mu, \sigma)$
$k_1 = 2$	Box-Cox Cole-Green	$\text{BCCG}(\mu, \sigma, \nu)$
	Lognormal	$\text{LOGNO}(\mu, \sigma)$
	Log Weibull	$\text{LOGWEI}(\mu, \sigma)$
$1 \leq k_1 < \infty$	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

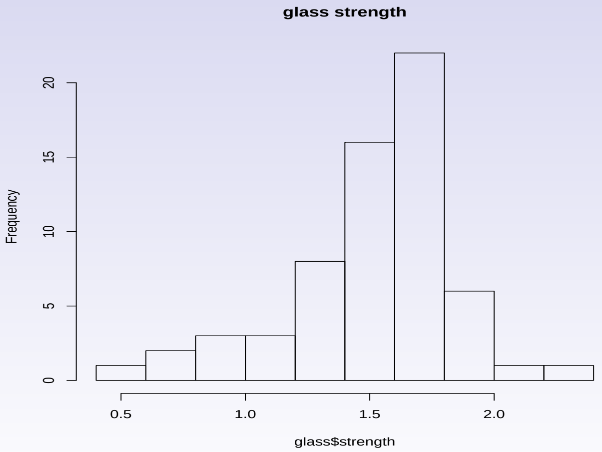
- optim() or mle() requires initial parameter values
- gamlsssML() mle a using variation of the mle() function
- gamlss(): mle using RS or CG or mixed algorithms,
- histdist() mle using RS or CG or mixed algorithms, for a univariate sample only, and plots a histogram of the data with the fitted distribution
- fitDist() fits a set of distributions and chooses the one with the smallest GAIC

Strength of glass fibres data

Data summary:

- R data file: glass in package `gamlss.data` of dimensions 63×1
- sourse: 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



gamlss

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

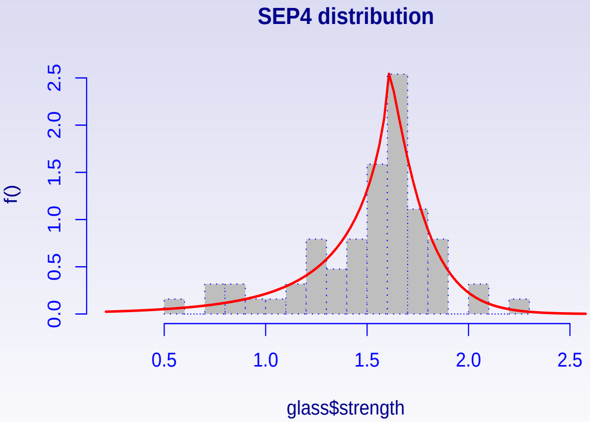
```
> m1<-fitDist(strength, data=glass, k=2, extra="TFtr")
        # AIC
> m2<-fitDist(strength, data=glass, k=log(length(strength))
         extra="TFtr") # SBC
> m1$fit[1:8]
    SEP4     SEP3   SHASHo     EGB2      JSU    BCPEo
27.68790 27.98191 28.01800 28.38691 30.41380 31.17693 31.40
> m2$fit[1:8]
    SEP4     SEP3   SHASHo     EGB2       GU     WEI3
36.26044 36.55444 36.59054 36.95945 38.19839 38.69995 38.98
```

gamlss

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



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 encounter when modelling count data using the Poisson distribution.

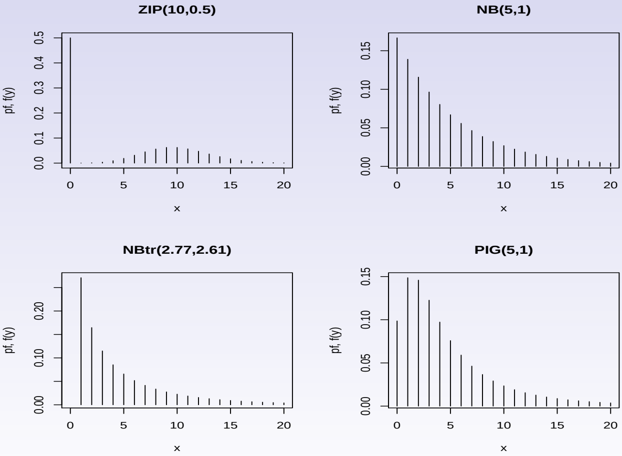
- overdispersion
- excess (or shortage) of zero values
- long tails (rare events)

Discrete distribution modelling

Par.	Modelling	Distributions
1	Location	PO
2	Location and scale	NBI, NBII, PIG
2	Location and zero probability	ZALG, ZAP, ZIP, ZIP2
3	Location, scale and skewness	DEL, SI, SICHEL
3	Location, scale and zero probability	ZANBI, ZINBI, ZIPIG

gamlss

Different count data distributions



gamlss

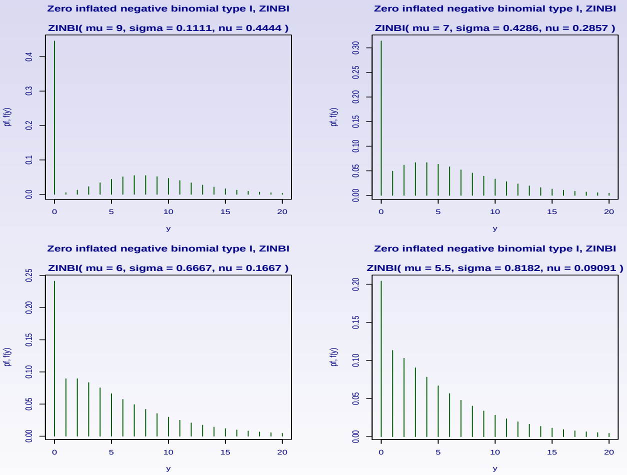
Zero inflated distributions

Zero inflated distribution, $Y \sim \textbf{ZID}$ is given by
$Y = 0$ with probability p
$Y \sim \textbf{D}$ with probability $1 - p$.
Hence

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

where $Y_1 \sim \textbf{D}$.

ZINBI distribution plots



Zero adjusted distributions

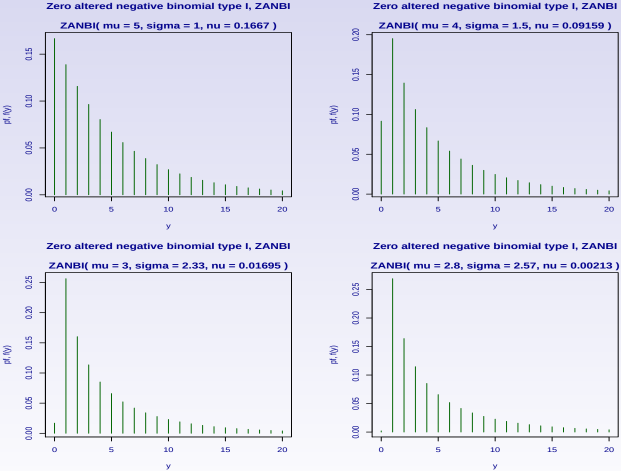
Zero adjusted distribution, $Y \sim \mathbf{ZAD}$ is given by
$Y = 0$ with probability p
$Y \sim \mathbf{Dtr}$ with probability $1 - p$,
where **Dtr** is a truncated distribution, **D** truncated at zero.
Hence

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

where $Y_1 \sim \mathbf{D}$.

gamlss

ZANBI distribution plots



gamlss

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 $\Re^+$ then $f_Y(y) = F_W(y+1) - F_W(y)$

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

gamlss

(c) Random effect at the observation level

(i) when an 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

$$\underbrace{f_Y(y)}_{\text{discrete}} = \int \underbrace{f(y|\gamma)}_{\text{discrete}} \underbrace{f_\gamma(\gamma)}_{\text{continuous}} \, d\gamma$$

- $Y \sim NBI(\mu, \sigma)$
- $Y|\gamma \sim PO(\gamma\mu)$
- $\gamma \sim GA(1, \sigma^{1/2})$

gamlss

Random effect at the observation level case (ii)

(ii) Non-explicit continuous mixture distribution

$$\underbrace{f_Y(y)}_{\text{discrete}} = \int \underbrace{f(y|\gamma)}_{\text{discrete}} \underbrace{f_\gamma(\gamma)}_{\text{continuous}} \; d\gamma$$

- $Y \sim PO - Normal(\mu, \sigma)$
- $Y|\gamma \sim PO(\gamma\mu)$
- $\log(\gamma) \sim NO(1, \sigma)$

Random effect at the observation level case (iii)

(iii) Non-parametric mixture distribution

$$\underbrace{f_Y(y)}_{\text{discrete}} = \sum_{k=1}^{K} \underbrace{f(y|\gamma_k)}_{\text{discrete}} \underbrace{p(\gamma = \gamma_k)}_{\text{continuous}}$$

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

where NPFM(2) equals Non-Parametric Finite Mixture with 2 point probabilities

gamlss

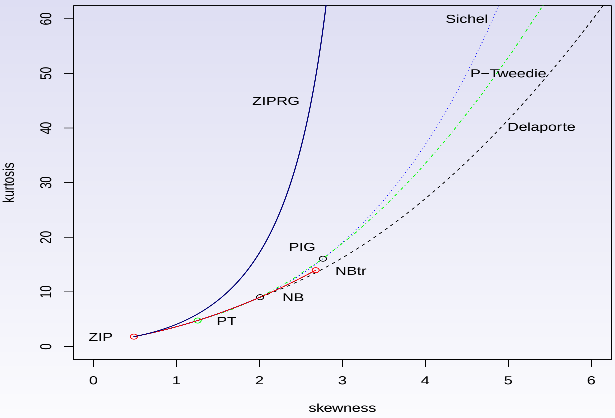
Explicit continuous mixture distribution

Distributions	R Name	mixing distribution for γ
Poisson	$\texttt{PO}(\mu)$	-
Neg. bin. I	$\texttt{NBI}(\mu, \sigma)$	$\texttt{GA}(1, \sigma^{\frac{1}{2}})$
Neg. bin. II	$\texttt{NBII}(\mu, \sigma)$	$\texttt{GA}(1, \sigma^{\frac{1}{2}}/\mu)$
Poisson IG	$\texttt{PIG}(\mu, \sigma)$	$\texttt{IG}(1, \sigma^{\frac{1}{2}})$
Sichel	$\texttt{SICHEL}(\mu, \sigma, \nu)$	$\texttt{GIG}(1, \sigma^{\frac{1}{2}}, \nu)$
Delaporte	$\texttt{DEL}(\mu, \sigma, \nu)$	$\texttt{SG}(1, \sigma^{\frac{1}{2}}, \nu)$
Zero inflated Poisson	$\texttt{ZIP}(\mu, \sigma)$	$\texttt{BI}(1, 1-\sigma)$
Zero inflated Poisson 2	$\texttt{ZIP2}(\mu, \sigma)$	$(1-\sigma)^{-1}\texttt{BI}(1, 1-\sigma)$
Zero inflated neg. bin.	$\texttt{ZINBI}(\mu, \sigma, \nu)$	zero inflated gamma
Poisson-Tweedie	-	Tweedie family

Table : Discrete gamlss family distributions for count data

R Name	params	mean	variance
$\mathtt{PO}(\mu)$	1	μ	μ
$\mathtt{NBI}(\mu, \sigma)$	2	μ	$\mu + \sigma\mu^2$
$\mathtt{NBII}(\mu, \sigma)$	2	μ	$\mu + \sigma\mu$
$\mathtt{PIG}(\mu, \sigma)$	2	μ	$\mu + \sigma\mu^2$
$\mathtt{SICHEL}(\mu, \sigma, \nu)$	3	μ	$\mu + h(\sigma, \nu)\mu^2$
$\mathtt{DEL}(\mu, \sigma, \nu)$	3	μ	$\mu + \sigma(1-\nu)^2\mu^2$
$\mathtt{ZIP}(\mu, \sigma)$	2	$(1-\sigma)\mu$	$(1-\sigma)\mu + \sigma(1-\sigma)\mu^2$
$\mathtt{ZIP2}(\mu, \sigma)$	2	μ	$\mu + \frac{\sigma}{(1-\sigma)}\mu^2$

Comparison of the marginal distributions using a (ratio moment) diagram of their skewness and kurtosis



gamlss

A stylometric application

Data summary:

R data file: `stylo` in package **gamlss.data** of dimensions 64×2

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.

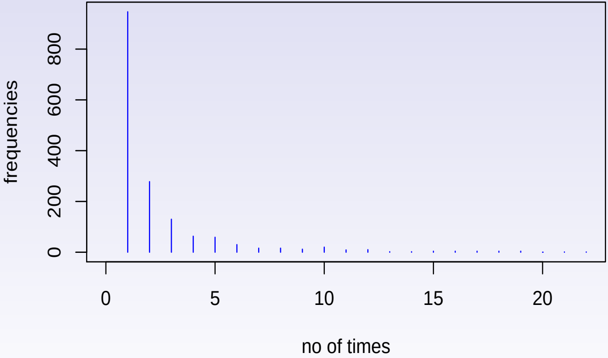
conclusion the truncated SICHEL distributions fits best

gamlss

A stylometric application

```
 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



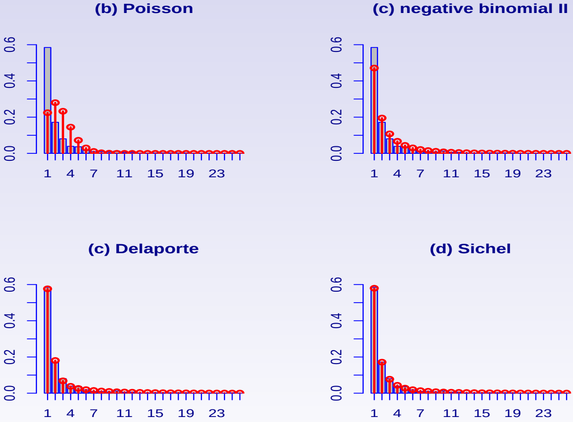
A stylometric application

```
> library(gamlss.tr)
> gen.trun(par = 0, family = PO, type = "left")
A truncated family of distributions from PO has been genera
 and saved under the names:
 dPOtr pPOtr qPOtr rPOtr POtr
The type of truncation is left and the truncation parameter
> 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 = POtr, trace = FALSE)
> mNBII <- gamlss(word ~ 1, weights = freq, data = stylo,
+ family = NBIItr, n.cyc = 50, trace = FALSE)
> mDEL <- gamlss(word ~ 1, weights = freq, data = stylo,
+ family = DELtr, n.cyc = 50, trace = FALSE)
> mSI <- gamlss(word ~ 1, weights = freq, data = stylo,
+ family = SICHELtr, n.cyc = 50, trace = FALSE)
> GAIC(mPO, mNBII, mDEL, mSI)
      df      AIC
mSI    3 5148.454
mDEL   3 5160.581
mNBII  2 5311.627
mPO    1 9207.459
```

The stylometric data



The fish species data

Data summary: the fish species data

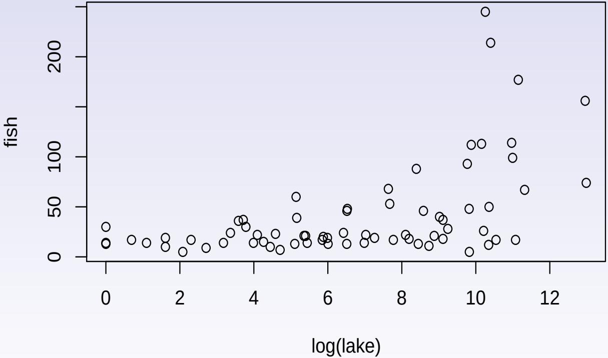
R data file: `species` in package gamlss.data of dimensions 70×2

variables

`fish` : the number of different species in 70 lakes in the world

`lake` : the lake area

The fish species data



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 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

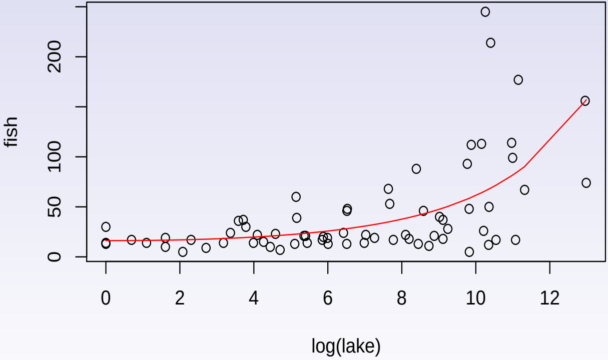
Table : Comparison of models for the fish species data

Model	$f_Y(y)$	μ	σ	ν	DEV	df	AIC	SBC
1	PO	$x < 2 >$	-	-	1849.3	3	1855.3	1862.0
2	NBI	x	1	-	619.8	3	625.8	632.6
3	NBI	$x < 2 >$	1	-	614.3	4	622.3	631.3
4	NBI	$cs(x, 3)$	1	-	611.9	6	623.9	637.4
5	NBI	$x < 2 >$	x	-	605.0	5	615.0	626.2
6	NBI-fam	$x < 2 >$	1	1	606.0	5	616.0	627.3
7	NBI-fam	$x < 2 >$	x	1	604.9	6	616.9	630.4

Overdispersed count data approaches

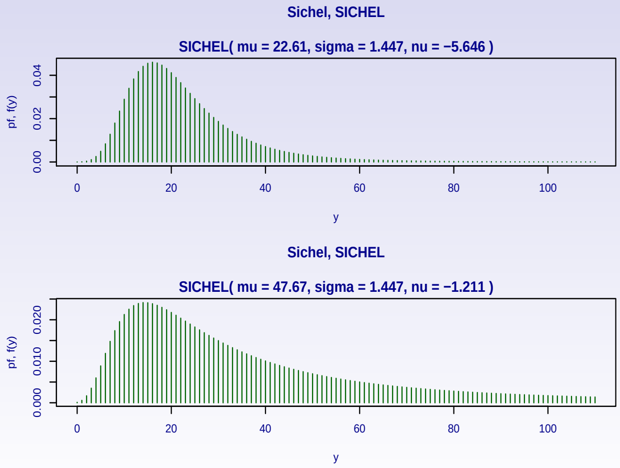
Model	$f_Y(y)$	μ	σ	ν	DEV	df	AIC	SBC
8	PIG	$x < 2 >$	1	-	613.3	4	621.3	630.3
9	SI	$x < 2 >$	1	x	597.7	6	609.7	623.2
10	DEL	$x < 2 >$	1	x	600.6	6	612.6	626.1
11	DEL	$x < 2 >$	-	x	600.6	5	610.6	621.9
12	PO-Normal	$x < 2 >$	1	-	615.2	4	623.2	632.2
13	NBI-Normal	$x < 2 >$	x	1	603.7	6	615.7	629.2
14	PO-NPFM(5)	$x < 2 >$	-	—	601.9	13	627.9	657.2
15	NB-NPFM(2)	$x < 2 >$	1	—	611.9	6	623.9	637.4
16	doublePO	$x < 2 >$	x	-	616.4	5	626.4	637.6
17	IGdisc	$x < 2 >$	1	-	603.3	4	611.3	620.3

Fitted mean of the Sichel distribution



gamlss

Fitted Sichel distributions for observations (a) 40 and (b) 67



Binomial response variables

There are only two distributions here

- binomial
- beta binomial

gamlss

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.

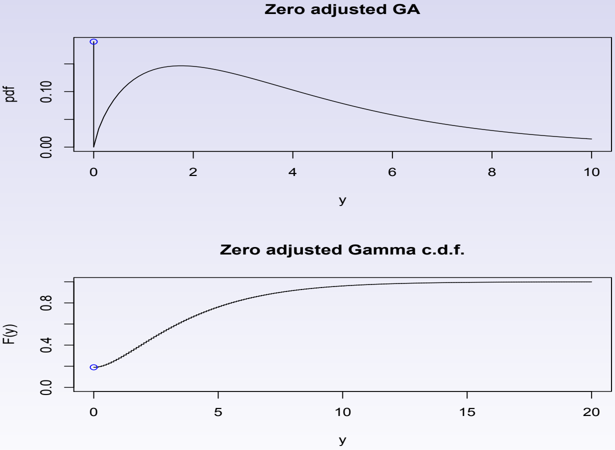
Zero adjusted distributions on zero and the positive real line $[0, \infty)$

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

$$f_Y(y) = \begin{cases} p & \text{if } y = 0 \\ (1-p) f_W(y) & \text{if } y > 0 \end{cases} \quad (2)$$

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 gamma distribution, **ZAGA**(μ,σ,ν)

The default link functions relating the parameters (μ, σ, ν) to the predictors (η_1, η_2, η_3), which may depend on explanatory variables, are

$$\log \mu = \eta_1$$

$$\log \sigma = \eta_2$$

$$\log \left(\frac{\nu}{1-\nu} \right) = \eta_3.$$

Zero adjusted gamma distribution, **ZAGA**(μ,σ,ν)

The ZAGA model is equivalent to

- a gamma distribution $GA(\mu, \sigma)$ model for $Y > 0$ together
- with a binary model for recoded variable Y_1 iven by

$$Y_1 = \begin{cases} 0 & \text{if } Y > 0 \\ 1 & \text{if } Y = 0 \end{cases} \tag{3}$$

i.e.

$$p(Y_1 = y_1) = \begin{cases} (1 - \nu) & \text{if } y_1 = 0 \\ \nu & \text{if } y_1 = 1 \end{cases} \tag{4}$$

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 $(1 - p_0 - p_1)$.

gamlss

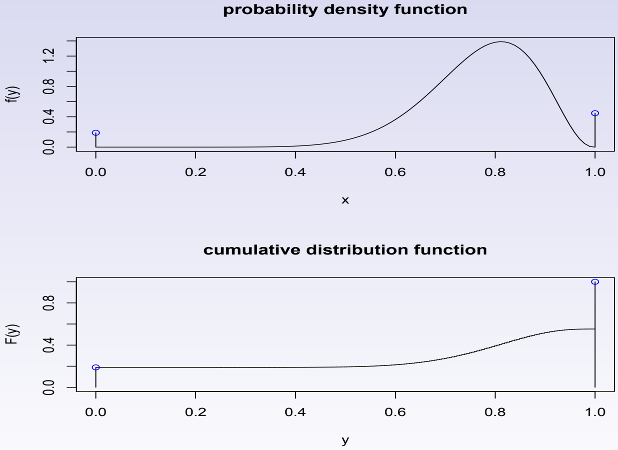
Distributions on the interval (0,1) inflated at 0 and 1

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 $0 \le y \le 1$, 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



END

for more information see

www.gamlss.org

gamlss