Flexible Regression and Smoothing

The gamlss.family Distributions

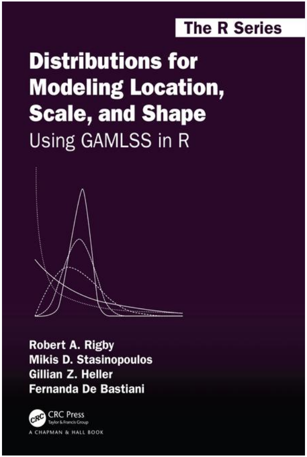
Mikis Stasinopoulos, Rob Rigby, Gilllian Heller,
and Fernanda De Bastiani

BCAM – Basque Center for Applied Mathematics, Bilbao, Basque Country - Spain, October 2019

gamlss

1. Properties of distributions
2. Types of distribution in GAMLSS
3. End

Information



gamlss

Response variable values

Continuous response:

- *the real line* $\Re = (-\infty, \infty)$
- *the positive real line* $\Re_+ = (0, \infty)$
- *values between zero and one* $\Re_0^1 = (0, 1)$.

Discrete response variable:

- the binary values : $\{0, 1\}$
- the binomial values : $\{0, 1, 2, \dots, \}$
- the non-negative integer values : $\{0, 1, 2, \dots\}$

gamlss

Response variable pdf

Continuous response:

$$\int_{R_Y} f(y)\, dy = 1 \ .$$

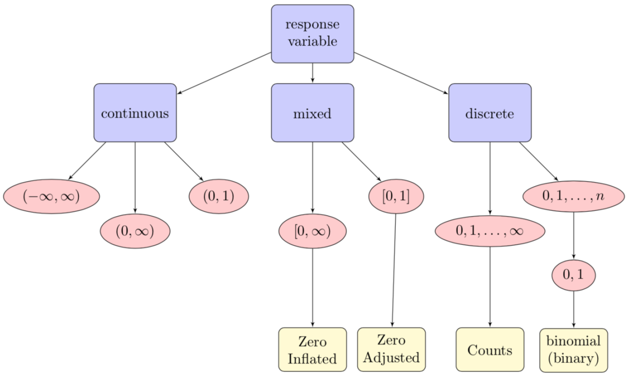
Discrete response variable:

$$\sum_{y \in R_Y} f(y) = \sum_{y \in R_Y} P(Y = y) = 1 \ .$$

Mixed response variable:

$$\int_{R_1} f(y)\, dy + \sum_{y \in R_2} f(y) = 1 \ .$$

Response variable selection of distribution

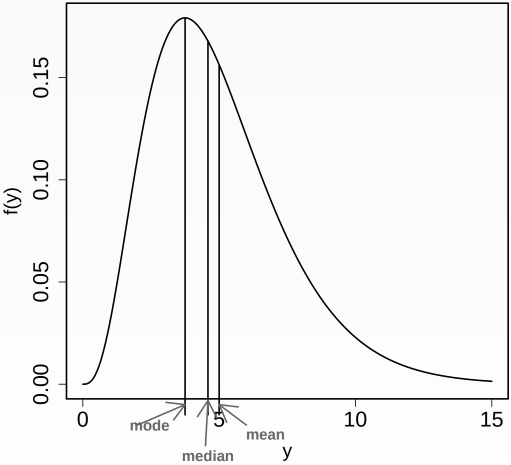


Properties

1. moment based properties;
2. centile based properties.

Location

- the mean, the average of Y, or expected value of Y, or
- the median, the value of Y which cuts the distribution in two halves, each with probability 0.50, or
- the mode, the value of Y which has the highest value of the probability (density) function.


f(y)
0.00
0.05
0.10
0.15
0
5
10
15
y
mode
median
mean

Scale

A scale parameter is related to the ‘spread’ of the distribution. Occasionally it is the:

- standard deviation or
- coefficient of variation

of the distribution.

Skewness

The skewness is a measure of distributional asymmetry.
Informally:

- a distribution with a heavier tail to the right than the left usually has positive skewness,
- one with a heavier tail to the left usually has negative skewness,
- a symmetric distribution has zero skewness.

see Chapter of 14 of Book 2, Rigby *et al.* (2019), for details

Kurtosis

The Kurtosis is a measure primary of heavy tails.
Informally:

- a distribution with heavy (i.e. fat) tails (relative to a normal distribution) will usually have leptokurtosis
- a distribution with light (i.e. thin) tails (relative to a normal distribution) will usually have platykurtosis.
- leptokurtic, and platykurtic distributions are judged by comparison to the normal distribution, which is called mesokurtic

see Chapter of 15 of Book 2, Rigby *et al.* (2019), for details

Skewness and kurtosis

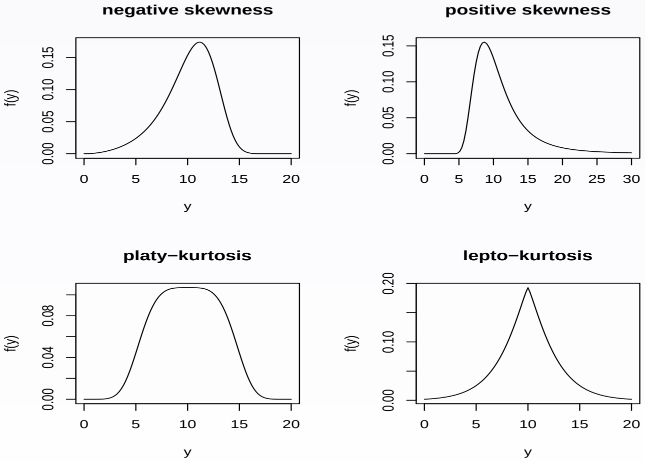


Figure: Showing different types of continuous distributions

Moment measures

mean

$$E(Y) = \begin{cases} \int_{-\infty}^{\infty} yf(y)\,dy & \text{for a continuous random variable } Y \\ \sum_{y \epsilon R_Y} y\,P(Y=y) & \text{for a discrete random variable } Y\ . \end{cases}$$

variance

$$V(Y) = \begin{cases} \int_{-\infty}^{\infty} [y - E(Y)]^2\, f(y)\,dy & \text{for continuous } Y \\ \sum_{y \in R_Y} [y - E(Y)]^2\, P(Y=y) & \text{for discrete } Y\ . \end{cases}$$

gamlss

Moment measures

moments about zero

$$\mu_k{}' = E\left(Y^k\right) \qquad \text{for } k = 1, 2, 3, \ldots$$

Hence $\mu_1{}' = E\left(Y\right)$.

central moments

$$\mu_k = E\left\{[Y - E(Y)]^k\right\} \qquad \text{for } k = 2, 3, \ldots \ .$$

Hence $\mu_2 = V(Y)$.

moment measure of skewness

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\mu_2{}^{1.5}} \ .$$

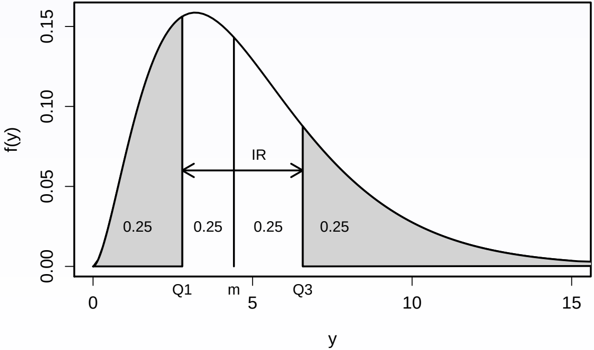
moment measure of kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2{}^2} \ .$$

gamlss

Centile measures

Centile measures always exist:



Centile measures of skewness

A centile skewness function of Y MacGillivray(1986) is:

$$s_p = s_p(Y) = \frac{(y_p + y_{1-p})/2 - y_{0.5}}{(y_{1-p} - y_p)/2} \qquad \text{for } 0 < p < 0.5 \;,$$

Galton's centile skewness

$$\gamma = \frac{(Q_3 + Q_1)/2 - m}{\text{SIR}} \;.$$

Centile measures kurtosis

A centile kutosis function of Y Balanda & MacGillivray (1988) is:

$$k_p = k_p(Y) = \frac{y_{1-p} - y_p}{Q_3 - Q_1} \qquad \text{for } 0 < p < 0.5 \ , \tag{1}$$

Andrew's centile kurtosis measure

$$\delta = \frac{y_{0.99} - y_{0.01}}{\text{IR}} \ .$$

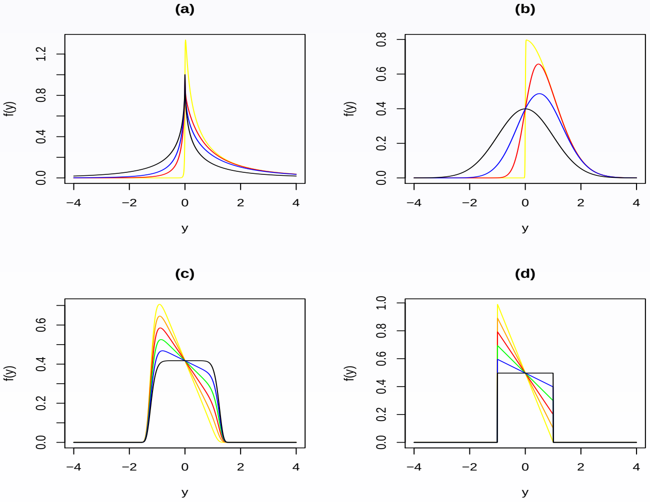
Different types of distribution in GAMLSS

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.

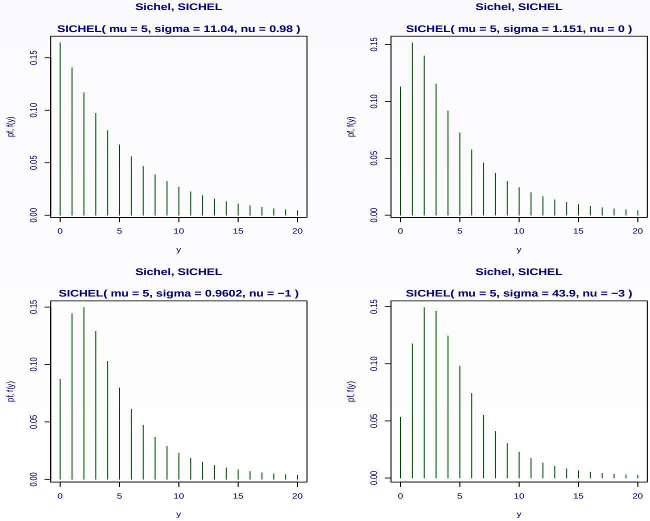
Different types of distribution: demos

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

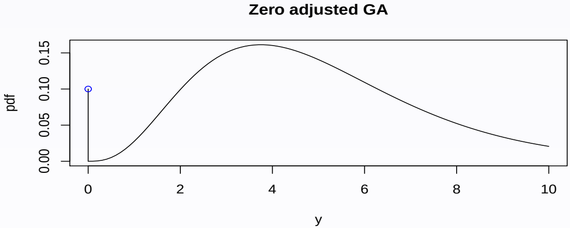
Example of continuous distribution: SEP1

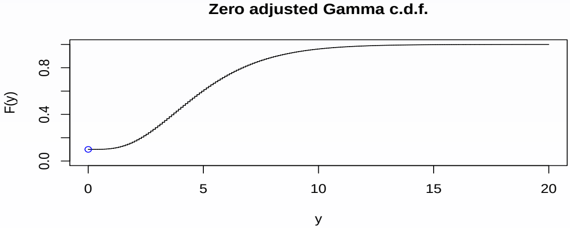


Example of discrete distribution: Sichel



Example of mixed distribution distributions: ZAGA





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)

END

for more information see

www.gamlss.com