

Flexible Regression and Smoothing

Mixed Distributions

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

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

- 1 Mixed distributions
 - Zero adjusted distributions: on zero and the positive real line $[0, \infty)$
 - Inflated distributions at 0 and 1
 - Data analysis: Lung function data

- 2 Finite mixtures

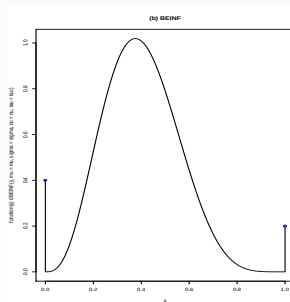
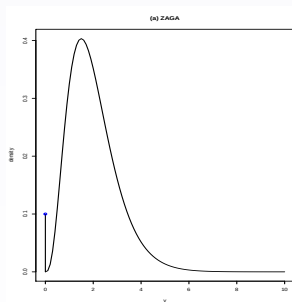
Mixed distributions

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.

Mixed distributions



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

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

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

Appropriate when $Y = 0$ has non-zero probability and otherwise $Y > 0$.
For example when Y

- measures the amount of rainfall in a day (where some days have zero rainfall),
- the river flow at a specific time each day (where some days the river flow is zero),
- the total amount of insurance claims in a year for individuals (where some people do not claim at all and therefore their total claim is zero).

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

The zero adjusted gamma distribution is a mixture of

- a discrete value 0 with probability ν , and
- a gamma $GA(\mu, \sigma)$ distribution on the positive real line $(0, \infty)$ with probability $(1 - \nu)$.

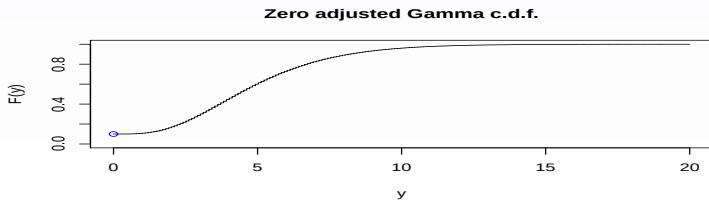
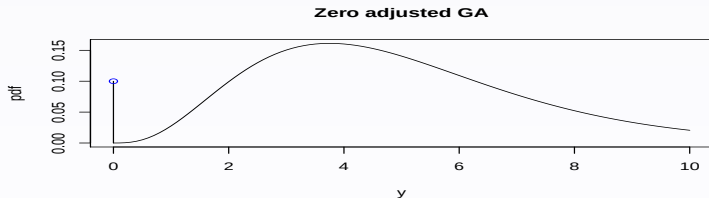
Zero adjusted gamma distribution, **ZAGA**(μ, σ, ν)

The probability (density) function of the zero adjusted gamma distribution, denoted by **ZAGA**(μ, σ, ν), is given by

$$f_Y(y|\mu, \sigma, \nu) = \begin{cases} \nu & \text{if } y = 0 \\ (1 - \nu)f_W(y|\mu, \sigma) & \text{if } y > 0 \end{cases} \quad (2)$$

for $0 \leq y < \infty$, where $\mu > 0$ and $\sigma > 0$ and $0 < \nu < 1$, and $W \sim GA(\mu, \sigma)$ has a gamma distribution.

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 given by

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

i.e.

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

Zero adjusted distributions

- package **gamlss.dist** explicit distribution: ZAGA() and ZAIG()
- package **gamlss.inf** generates distributions using the function `gen.Zadj()` which takes arguments any 0 to ∞ distribution:

See vignette:

"Zero adjusted distributions on the positive real line"

Zero adjusted distributions

The GAMLSS model changes to

$$Y \stackrel{\text{ind}}{\sim} \mathcal{D}(\mu, \sigma, \nu, \tau, \xi_0)$$

$$\eta_1 = g_1(\mu) = \mathbf{X}_1\beta_1 + s_{11}(\mathbf{x}_{11}) + \dots + s_{1J_1}(\mathbf{x}_{1J_1})$$

$$\eta_2 = g_2(\sigma) = \mathbf{X}_2\beta_2 + s_{21}(\mathbf{x}_{21}) + \dots + s_{2J_2}(\mathbf{x}_{2J_2})$$

$$\eta_3 = g_3(\nu) = \mathbf{X}_3\beta_3 + s_{31}(\mathbf{x}_{31}) + \dots + s_{3J_3}(\mathbf{x}_{3J_3})$$

$$\eta_4 = g_4(\tau) = \mathbf{X}_4\beta_4 + s_{41}(\mathbf{x}_{41}) + \dots + s_{4J_4}(\mathbf{x}_{4J_4})$$

$$\eta_5 = g_5(\xi_0) = \mathbf{X}_5\beta_5 + s_{51}(\mathbf{x}_{51}) + \dots + s_{5J_5}(\mathbf{x}_{5J_5})$$

Note that ξ_0 is the probability at zero

Inflated distributions 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 $(1 - p_0 - p_1)$.

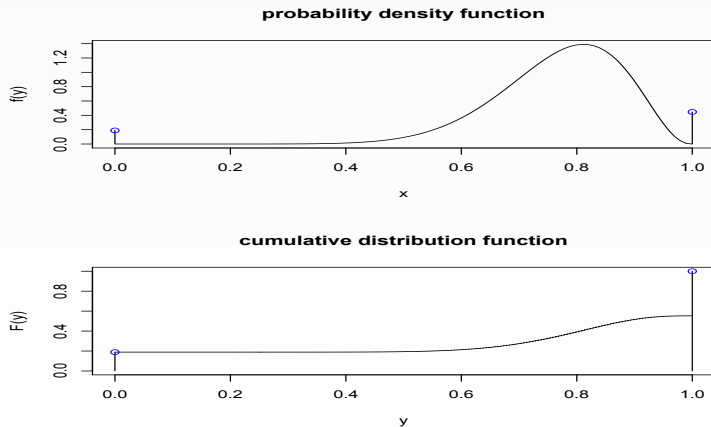
Inflated distributions at 0 and 1

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

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

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



Beta inflated distribution, **BEINF**(μ, σ, ν, τ)

The probability (density) function of the beta inflated distribution, denoted by **BEINF**(μ, σ, ν, τ), is defined by

$$f_Y(y|\mu, \sigma, \nu, \tau) = \begin{cases} p_0 & \text{if } y = 0 \\ (1 - p_0 - p_1)f_W(y|\mu, \sigma) & \text{if } 0 < y < 1 \\ p_1 & \text{if } y = 1 \end{cases} \quad (6)$$

for $0 \leq y \leq 1$, where $W \sim BE(\mu, \sigma)$ has a beta distribution, $\nu = p_0/p_2$ and $\tau = p_1/p_2$ where $p_2 = 1 - p_0 - p_1$.

BEINF: demo

1 demo.BEINF()

The default link functions relating the parameters (μ, σ, ν, τ) to the predictors $(\eta_1, \eta_2, \eta_3, \eta_4)$, which may depend on explanatory variables, are

$$\log \left(\frac{\mu}{1 - \mu} \right) = \eta_1$$

$$\log \left(\frac{\sigma}{1 - \sigma} \right) = \eta_2$$

$$\log \nu = \log \left(\frac{p_0}{p_2} \right) = \eta_3$$

$$\log \tau = \log \left(\frac{p_1}{p_2} \right) = \eta_4$$

The model is equivalent to

- a beta distribution $BE(\mu, \sigma)$ model for $0 < Y < 1$
- a multinomial model $MN3(\nu, \tau)$ with three levels for recoded variable Y_1 given by

$$Y_1 = \begin{cases} 0 & \text{if } Y = 0 \\ 1 & \text{if } Y = 1 \\ 2 & \text{if } 0 < Y < 1 \end{cases} \quad (7)$$

i.e.

$$p(Y_1 = y_1) = \begin{cases} p_0 & \text{if } y_1 = 0 \\ p_1 & \text{if } y_1 = 1 \\ 1 - p_0 - p_1 & \text{if } y_1 = 2 \end{cases} \quad (8)$$

where $\nu = p_0/p_2$ and $\tau = p_1/p_2$ where $p_2 = 1 - p_0 - p_1$

Beta inflated at 0 distribution, **BEINF0**(μ, σ, ν)

The probability (density) function of the beta inflated at 0 distribution, denoted by **BEINF0**(μ, σ, ν), is given by

$$f_Y(y|\mu, \sigma, \nu) = \begin{cases} p_0 & \text{if } y = 0 \\ (1 - p_0)f_W(y|\mu, \sigma) & \text{if } 0 < y < 1 \end{cases} \quad (9)$$

for $0 \leq y < 1$, where $W \sim BE(\mu, \sigma)$ has a beta distribution where $\nu = p_0/(1 - p_0)$.

Beta inflated at 1 distribution, **BEINF1**(μ, σ, ν)

The probability (density) function of the beta inflated at 1 distribution, denoted by **BEINF1**(μ, σ, ν), is given by

$$f_Y(y|\mu, \sigma, \nu) = \begin{cases} p_1 & \text{if } y = 1 \\ (1 - p_1)f_W(y|\mu, \sigma) & \text{if } 0 < y < 1 \end{cases} \quad (10)$$

for $0 < y \leq 1$, where $W \sim BE(\mu, \sigma)$ has a beta distribution where $\nu = p_1/(1 - p_1)$.

Inflated distributions

- package **`gamlss.dist`** implicit distributions: `BEINF()` and `BEINF0()`, `BEZI()`, `BEINF1()` `BEOI()`
- package **`gamlss.inf`** generates distributions using the function `gen.Inf0to1()`
which takes arguments any
 - explicit distributions defined on $(0, 1)$
 - transformed distributions defined on $(0, 1)$, i.e `gen.Family( ..., "logit")`
 - truncated distributions defined on $(0, 1)$ i.e `gen.trun(c(0,1), "SST", type="both")`

See vignette:

"Inflated distributions on the interval $[0,1]$ "

Inflated distributions

$$Y \stackrel{\text{ind}}{\sim} \mathcal{D}(\mu, \sigma, \nu, \tau, \xi_0, \xi_1)$$

$$\eta_1 = g_1(\mu) = \mathbf{X}_1\beta_1 + s_{11}(\mathbf{x}_{11}) + \dots + s_{1J_1}(\mathbf{x}_{1J_1})$$

$$\eta_2 = g_2(\sigma) = \mathbf{X}_2\beta_2 + s_{21}(\mathbf{x}_{21}) + \dots + s_{2J_2}(\mathbf{x}_{2J_2})$$

$$\eta_3 = g_3(\nu) = \mathbf{X}_3\beta_3 + s_{31}(\mathbf{x}_{31}) + \dots + s_{3J_3}(\mathbf{x}_{3J_3})$$

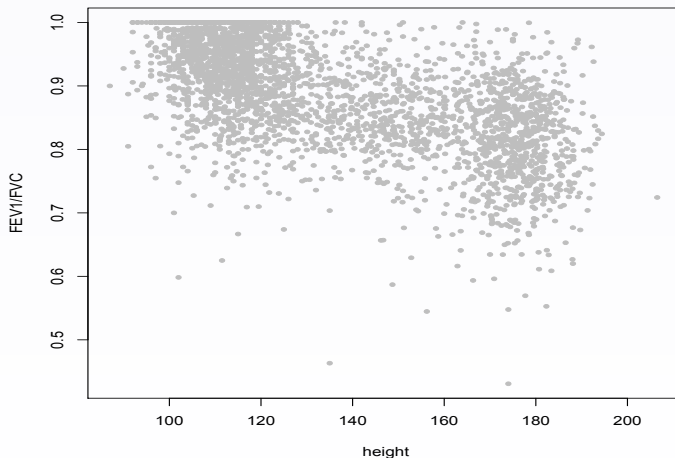
$$\eta_4 = g_4(\tau) = \mathbf{X}_4\beta_4 + s_{41}(\mathbf{x}_{41}) + \dots + s_{4J_4}(\mathbf{x}_{4J_4})$$

$$\eta_5 = g_5(\xi_0) = \mathbf{X}_5\beta_5 + s_{51}(\mathbf{x}_{51}) + \dots + s_{5J_5}(\mathbf{x}_{5J_5})$$

$$\eta_6 = g_6(\xi_1) = \mathbf{X}_6\beta_6 + s_{61}(\mathbf{x}_{61}) + \dots + s_{6J_6}(\mathbf{x}_{6J_6})$$

Note that ξ_0 and ξ_1 are related to the probability at zero and one

3164 male observations of lung function data



The lung function data

$Y = FEV_1/FVC$: the Spirometric lung function an established index for diagnosing airway obstruction (3164 male)

age : the height in cm

Source: Stanojevic et al. 2009

Lung function data: distribution

A logitSST distribution inflated at 1 was used. The probability (density) function of Y is $f_Y(y)$ given by

$$f_Y(y|\mu, \sigma, \nu, \tau, p) = \begin{cases} p & \text{if } y = 1 \\ (1-p)f_W(y|\mu, \sigma, \nu, \tau) & \text{if } 0 < y < 1 \end{cases} \quad (11)$$

for $0 < y \leq 1$, where $W \sim \text{logitSST}(\mu, \sigma, \nu, \tau)$ has a logitSST distribution with $-\infty < \mu < \infty$ and $\sigma > 0$, $\nu > 0$, $\tau > 0$ and where $0 < p < 1$.

Lung function data: link functions

The default link functions relate the parameters $(\mu, \sigma, \nu, \tau, p)$ to the predictors $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)$, which are modelled as smooth functions of $lht = \log(\text{height})$, i.e.

$$\mu = \eta_1 = s(lht)$$

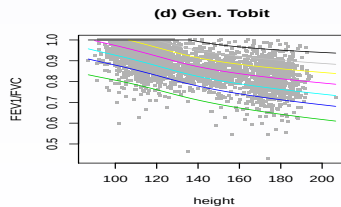
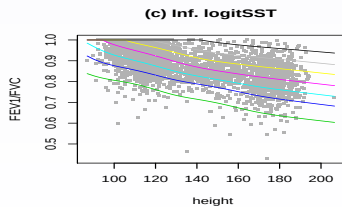
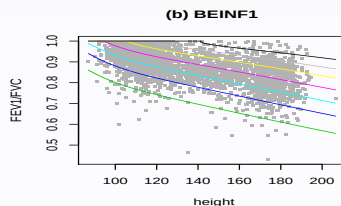
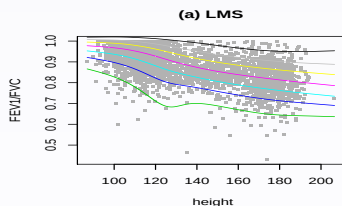
$$\log \sigma = \eta_2 = s(lht)$$

$$\log \nu = \eta_3 = s(lht)$$

$$\log \tau = \eta_4 = s(lht)$$

$$\log \left(\frac{p}{1-p} \right) = \eta_5 = s(lht)$$

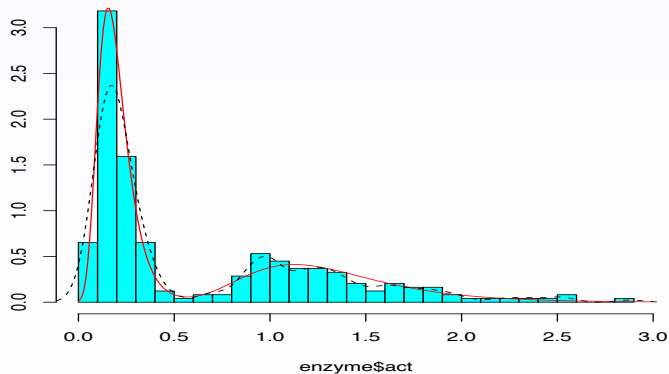
The lung function data: fitted centile curves



Finite mixtures: Why

- Dealing with **multimodal** distributions
- A way to introduce simple random effect models in GAMLSS
- distinction between
 - Finite mixtures with **no** parameters in common
 - Finite mixtures **with** parameters in common

Finite mixtures example: Enzyme data



Finite mixtures: Distribution function

Suppose that the random variable Y comes from component k , having probability (density) function $f_k(y)$, with probability π_k for $k = 1, 2, \dots, K$, then the (marginal) density of Y is given by

$$f_Y(y) = \sum_{k=1}^K \pi_k f_k(y)$$

where $0 \leq \pi_k \leq 1$ is the prior (or mixing) probability of component k , for $k = 1, 2, \dots, K$ and $\sum_{k=1}^K \pi_k = 1$.

Finite mixtures: Distribution function

The probability (density) function $f_k(y)$ for component k may depend on parameters θ_k and explanatory variables $\mathbf{x}_k$, i.e. $f_k(y) = f_k(y|\theta_k, \mathbf{x}_k)$. Similarly $f_Y(y)$ depends on parameters $\psi = (\theta, \pi)$ where $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ and $\pi^T = (\pi_1, \pi_2, \dots, \pi_K)$ and explanatory variables $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K)$, i.e. $f_Y(y) = f_Y(y|\psi, \mathbf{x})$, and

$$f_Y(y|\psi, \mathbf{x}) = \sum_{k=1}^K \pi_k f_k(y|\theta_k, \mathbf{x}_k)$$

Finite mixtures: the log Likelihood

$$\ell = \ell(\boldsymbol{\psi}, \mathbf{y}) = \sum_{i=1}^n \log \left[\sum_{k=1}^K \pi_k f_k(y_i) \right]$$

We wish to maximize ℓ with respect to $\boldsymbol{\psi}$, i.e. with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. It turns out that it is easier to maximise using **EM** algorithm:

- define the full likelihood
- take expectations
- maximise

Finite mixtures: the complete log Likelihood

$$\delta_{ik} = \begin{cases} 1, & \text{if observation } i \text{ comes from component } k \\ 0, & \text{otherwise} \end{cases}$$

Let $\delta_i^T = (\delta_{i1}, \delta_{i2}, \dots, \delta_{ik})$ be the indicator vector for observation i .

Let $\delta^T = (\delta_1^T, \delta_2^T, \dots, \delta_n^T)$ combine all the indicator variable vectors.

$$\ell_c = \ell_c(\psi, \mathbf{y}, \delta) = \sum_{i=1}^n \sum_{k=1}^K \delta_{ik} \log f_k(y_i) + \sum_{i=1}^n \sum_{k=1}^K \delta_{ik} \log \pi_k$$

Finite mixtures: EM-steps

E-step

$$\begin{aligned} Q &= E_{\boldsymbol{\delta}} \left[\ell_c | \mathbf{y}, \hat{\boldsymbol{\psi}}^{(r)} \right] \\ &= \sum_{k=1}^K \sum_{i=1}^n \hat{w}_{ik}^{(r+1)} \log f_k(y_i) + \sum_{k=1}^K \sum_{i=1}^n \hat{w}_{ik}^{(r+1)} \log \pi_k \end{aligned}$$

M-step weighted log likelihood for GAMLSS model

Finite mixtures: the weights

$$\begin{aligned}\hat{w}_{ik}^{(r+1)} &= E \left[\delta_{ik} | \mathbf{y}, \hat{\boldsymbol{\psi}}^{(r)} \right] \\ &= \frac{\hat{\pi}_k^{(r)} f_k(y_i | \hat{\boldsymbol{\theta}}_k^{(r)})}{\sum_{k=1}^K \hat{\pi}_k^{(r)} f_k(y_i | \hat{\boldsymbol{\theta}}_k^{(r)})}\end{aligned}$$

Finite mixtures: the `gamlssMX()` function

```
m1 <- gamlssMX(act ~ 1, family = NO, K = 2)
m2 <- gamlssMX(act ~ 1, family = GA, K = 2)
m3 <- gamlssMX(act ~ 1, family = RG, K = 2)
m4 <- gamlssMX(act ~ 1, family = c(NO, GA), K = 2)
m5 <- gamlssMX(act ~ 1, family = c(GA, RG), K = 2)
```

```
AIC(m1, m2, m3, m4, m5)
```

	df	AIC
m3	5	96.29161
m5	5	101.04612
m2	5	102.42911
m4	5	112.89527
m1	5	119.28005

.....

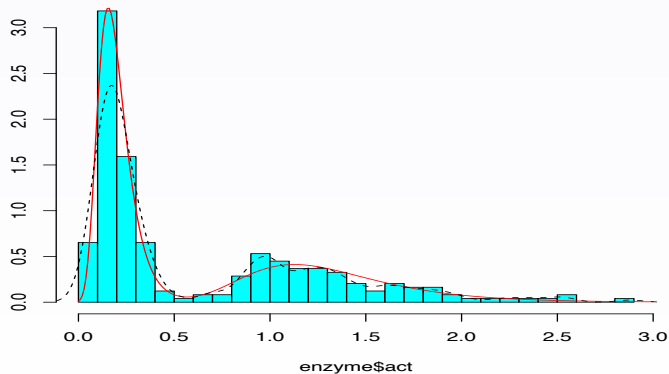
Finite mixtures: the `gamlssMX()` function

```
> m3  
Mu Coefficients for model: 1  
(Intercept)  
1.127  
Sigma Coefficients for model: 1  
(Intercept)  
-1.091  
Mu Coefficients for model: 2  
(Intercept)  
0.1557  
Sigma Coefficients for model: 2  
(Intercept)  
-2.641  
Estimated probabilities: 0.3760177 0.6239823
```

Finite mixtures: the `gamlssMX()` function

```
truehist(enzyme$act, h = 0.1)
fyRG <- dMX(y = seq(0, 3, 0.01),
  mu = list( 1.127, 0.1557),
  sigma = list(0.336, 0.0713),
  pi = list(0.376, 0.624),
  family = list("RG","RG"))
lines(seq(0, 3, 0.01), fyRG, col = "red", lty = 1)
lines(density(enzyme$act, width = "SJ-dpi"), lty = 2)
```


Finite mixtures example: Enzyme data



Finite mixtures: conclusions

- Finite mixtures of K components, each having a GAMLSS model, can be fitted using `gamlssMX()` if the K components have no parameters in common
- Modelling the mixing probabilities can be done (a multinomial logistic model is used)
- Finite mixtures with parameters in common can be fitted using the function `gamlssNP()`
- Mixed distributions are special case of finite mixtures

END

for more information see

www.gamlss.com