Illustration of the robust fitting using the `gamlssRobust()` function

F. Marmolejo-Ramos, R. Ospina, and F. Hernández-Barajas

June 28, 2022

In this document we present two examples of using the `gamlssRobust()` function to fit gamlss models when the response variable Y is contaminated with outliers.

In the first example we explore the robust gamlss fitting procedure using a mixture distribution with bounded data. The underlying distribution of the simulated data is the Beta distribution $Beta(\mu = 5, \sigma = 5)$ with symmetric shape around the value 0.5. However, we induce a two-side contamination using data from a uniform distribution $\mathcal{U}(a, b)$. The data used in this example are summarized as follows:

$$y \sim \alpha_1 \mathcal{U}(0, 0.1) + \alpha_2 \mathcal{U}(0.9, 1) + (1 - \alpha_1 - \alpha_2) Beta(5, 5). \tag{1}$$

Here $\alpha_1 = 0.01 = 1\%$ (contamination to lower-tail), $\alpha_2 = 0.01 = 1\%$ (contamination to upper-tail).

In the second example we explore the robust gamlss fitting procedure using a mixture distribution with positive data. The underlying distribution of the simulated data is the $ExGaussian(\mu = 0.5, \sigma = 0.1, \nu = 3)$ distribution with $\alpha = 0.1 = 10\%$ of contamination using data from a uniform distribution $\mathcal{U}(a, b)$. The data used in this example are summarized as follows:

$$y \sim (1 - \alpha) \, ExGaussian(0.5, 0.1, 3) + \alpha \mathcal{U}(0, 10). \tag{2}$$

Figures 1 and 2 show the simulated data by highlighting the induced outliers with red stars.

Using the simulated data shown in Figure 1, we fitted there models: Beta, Beta with bias correction and robust fitting model. Table 1 contains the estimated parameters for μ and σ, the last line shows the true values for those parameters. From this table we note that the estimations for the robust fitting model are closer to the true values $\mu = 5$ and $\sigma = 5$. Additionally, the robust fitting model presents the lower AIC (Akaike Information Criterion) value.

Using the simulated data shown in Figure 2, we fitted two models: ExGaussian and robust beta model. Table 2 contains the estimated parameters for μ, σ and ν, the last line shows the true values for those parameters. From this table we note that the estimations for the robust fitting model are closer to the true values $\mu = 0.5$, $\sigma = 0.1$ and $\nu = 3$.

Figure 3 shows three plots to explore the results in a graphical way. The firts row contains the results for the Beta example and the second row the results for the ExGaussian example. Firts we analize the results for the Beta example shown in first row of the figure. In the left panel there are the estimated density $f(y)$ curves under Beta, Beta with bias correction and robust fitting model, it is clear that green curve explains better the observed pattern in the histogram and this curve is closer
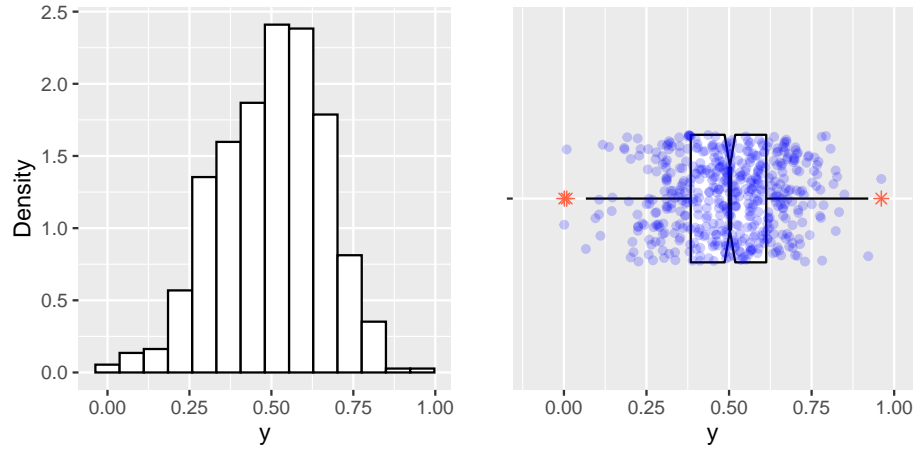
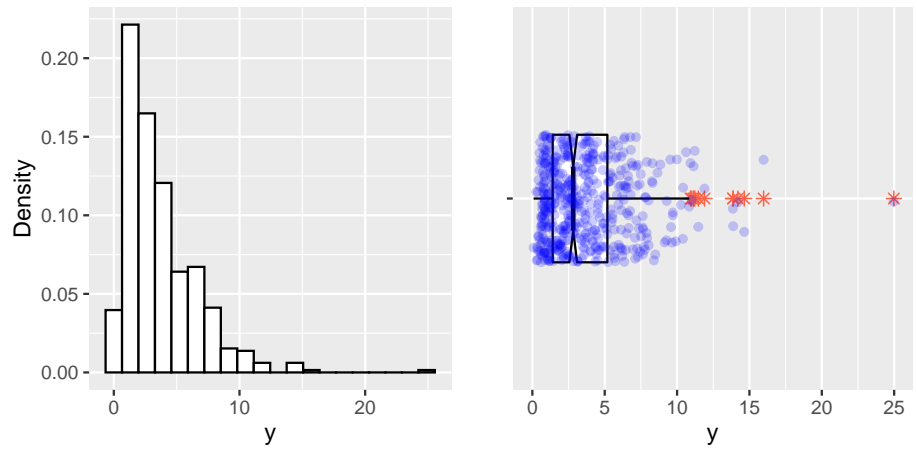Figure 1: Histogram (left panel) and boxplot (rigth panel) for $n = 500$ simulated data from model (1).

Figure 2: Histogram (left panel) and boxplot (rigth panel) for $n = 500$ simulated data from model (2).

Model	μ	σ	AIC
Beta	3.90	4.00	-376.35
Beta with bias correction	4.64	4.67	-368.22
Robust Beta	4.89	4.91	-462.07
True	5.00	5.00	–

Table 1: Estimated values for μ and σ under Beta, Beta with bias correction and robust fitting model. Last column presents the AIC (Akaike Information Criterion) and last line contains the true values for μ and σ parameters.

Model	μ	σ	ν	AIC
ExGaussian	0.48	0.18	3.16	2208.98
Robust ExGaussian	0.48	0.15	3.12	2179.49
True	0.50	0.10	3.00	–

Table 2: Estimated values for μ, σ and ν under ExGaussian and Robust ExGaussian. Last column presents the AIC (Akaike Information Criterion) and last line contains the true values for μ, σ and ν parameters.

to the the gray line that represents the true density. In the middle panel we observe the estimated cumulative $F(y)$ curves under the fitted models with the empirical cumulative density function represented by small gray dots. With careful observation, we can notice that the red curve moves away from the empirical cumulative density function. In the right panel there are the estimated hazard function $h(y)$ under Beta, Beta with bias correction and robust fitting models. From this plot we note that green and gray curves overlap each other. From this figure we can conclude that the robust fitting explains better the pattern for the simulated and contaminated data. Similar patterns are observe for the ExGaussian example shown in the second row of the Figure 3.

Figure 4 shows the worm plot proposed by Buuren and Fredriks (2001) which provides a guide to the adequacy of the fitted model. The worm plots for the robust fitting show a better pattern indicating that the robust fitting model is adequate for the data in both examples, Beta and ExGaussian.

The R codes to replicate both examples presented in this document can be found in a github repository that can be found in this url `https://github.com/fhernanb/dist_gamlss_book/tree/main/%5B03%5D%20Robust%20estimation%20example`.

References

Buuren, S. v., & Fredriks, M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in medicine*, *20*(8), 1259–1277.
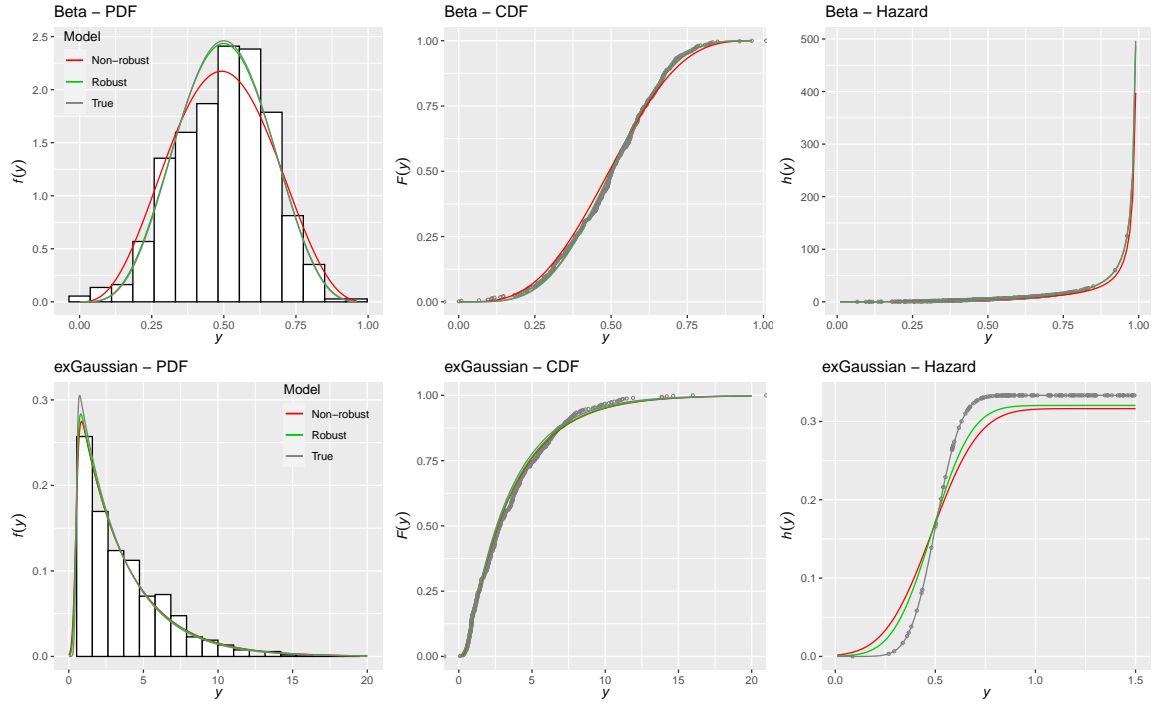
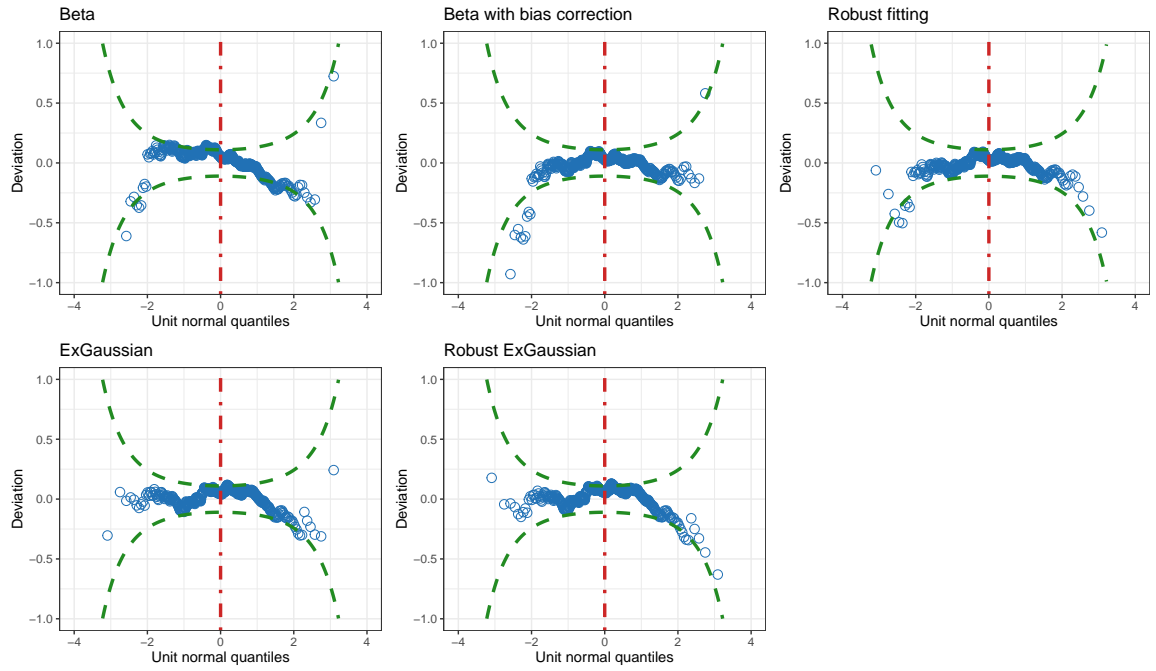Figure 3: Estimated $f(y)$, $F(y)$ and $h(y)$ for Beta and exGaussian examples.

Figure 4: Worm plots for Beta and exGaussian examples.