The place of probability distributions in statistical learning. A commented book review of "Distributions for modeling location, scale, and shape using GAMLSS in R" by Rigby et al. (2021). =*Supplementary file*=

F. Marmolejo-Ramos [1*], R. Ospina [2], & F. Hernández Barajas [3]

[1] *Center for Change and Complexity in Learning, University of South Australia, Adelaide, Australia.*
[2] *Departamento de Estatística, CASTLab, Universidade Federal de Pernambuco, Recife, Brazil.*
[3] *Escuela de Estadística, Universidad Nacional de Colombia, sede Medellín, Colombia.*

Summary

This section presents a summary of the book.

Key words: GAMLSS; probability distributions; distributional regression; robustness; skewness; kurtosis.

1. Probability distributions usable within GAMLSS

The book "Distributions for modelling location, scale, and shape using GAMLSS in R" (Rigby et al. 2021) is a companion to the book "Flexible regression and smoothing using GAMLSS in R" (Stasinopoulos et al. 2017). While the book by Stasinopoulos et al. (2017) presents the cogs of GAMLSS modelling (implemented in the **gamlss** R package), the book by Rigby et al. (2021) focuses on all probability distributions that can be used in GAMLSS modelling (implemented in the **gamlss.dist** R package). Also, as in their 2017 book, Rigby et al. (2021)'s book ends each Chapter with bibliographic notes that give extra references and historical context for some distributions and exercises for the reader

*Author to whom correspondence should be addressed.
[1] Center for Change and Complexity in Learning, University of South Australia, 61-68 North Terrace, Adelaide SA 5000, Australia. Email: `fernando.marmolejo-ramos@unisa.edu.au` (R.O.: `raydonal@de.ufpe.br` and F.H-B.: `fhernanb@unal.edu.co`)
Acknowledgment. The authors gratefully acknowledge feedback from Fernanda de Bastiani and Thomas Kneib. R codes and data sets used for the figures shown in this review are available at `https://cutt.ly/mGRadjE` (other examples and handy GAMLSS-related R functions are also available therein). We thank Mikis Stasinopolous for sharing the R code for robust GAMLSS.

(note the `gamlss.data` R package has all the data sets used in Rigby et al. (2021) and Stasinopoulos et al. (2017)). A brief tutorial on GAMLSS can be found in Stasinopoulos, Rigby & de Bastiani (2018) and a critical review of the book by Stasinopoulos et al. (2017) can be found in Welsh (2019).

Rigby et al. (2021)'s book has three parts: Part I ("Parametric distributions and the GAMLSS family of distributions") presents several parametric distributions and their properties. All are implemented in R and are usable within GAMLSS. The continuous distributions covered have support in the real line, the positive real line, and between zero and one and can take non-symmetric shapes such that they have positive to negative skewness and high to low kurtosis. An example of this type of distribution is the Ex-Gaussian; a distribution commonly used in cognitive neuroscience to model reaction times (Marmolejo-Ramos et al. in press). Discrete distributions and mixtures of continuous and discrete (mixed) distributions are also considered. This last type of distribution, although commonplace in practice, is rather ignored by applied researchers. For example, visual analogue scales (VAS) are primarily used in psychometric research. Although the range of the VAS's response variable is $[0, 1]$ (i.e. it is an inflated distribution), most researchers are unaware that GAMLSS and continuous ordinal regression (Manuguerra, Heller & M. 2020) are more appropriate techniques than classic linear regression to model such type of data. This part of the book is in line with that by Krishnamoorthy (2016) and that also covers several distributions (it also includes R codes).

Part II ("Advanced topics") contains eight Chapters. Chapter 10 ("Statistical inference") covers topics such as statistical inference, likelihood function, and frequentist and Bayesian approaches to inference. Chapter 11 ("Maximum likelihood estimation") discusses maximum likelihood estimators (MLE) of the parameters, and MLE's properties and the estimation of standard errors, confidence intervals, and tests. Chapter 12 ("Robustness of parameter estimation to outlier contamination") covers aspects relating to the robust fitting of GAMLSS models. Chapter 13 ("Methods of generating distributions") explains how new continuous distributions can be generated, how some interrelate, and families of distributions. The subsequent Chapters discuss skewness (Chapter 14), kurtosis (Chapter 15), how these compare in the case of continuous distributions (Chapter 16), and the tail heaviness of discrete and continuous distributions (Chapter 17).

Finally, Part III ("Reference guide") is essentially a 'look-up' version of Part I in that it provides the key properties of all distributions covered in the book and that are implemented in the **gamlss.dist** R package. The authors summarise via tables the essential properties of all distributions such as their ranges of the response variable and the parameters of the distribution, distribution measures (e.g. mean, median, etc), moment or probability generating functions, probability density or mass functions, cumulative distribution functions, inverse cumulative functions, and references to publications dealing with the distribution of interest. R code for most of the plots shown in this part of the book is provided in the book's appendix ("R code for plots of the reference guide"). It is important to note that distributions not implemented in **gamlss.dist** R package can still be used with GAMLSS so long they are coded appropriately (see section 6.4 in Chapter 6 in Stasinopoulos et al. (2017)). For example, the Birnbaum-Saunders distribution is not included in the **gamlss.dist** R package but Roquim et al. (2021) implemented it outside this package so that it can be used for GAMLSS modelling. A reparameterised version of that distribution is also usable via GAMLSS through the **rbsmodels** R package developed by Manoel Santos-Neto. The **RelDists** R package features more than 50 distributions usable with GAMLSS (this package is being maintained by Freddy Hernández-Barajas and colleagues and can be found at `https://ousuga.github.io/RelDists/`). Other R packages that contain other probability distributions and supporting estimating functions are **ssdtools** (by Joe Thorley and colleagues) and **fitdistrplus** (by Marie-Laure Delignette-Muller and Christophe Dutang).

While Part I and Part III are indeed relevant to the purpose of the book, Part II is perhaps the most exciting section. It deals with topics that link the GAMLSS framework and probability distributions to 'hot' topics in statistical learning (Gelman & Vehtari 2021; Friedrich et al. 2021). For example, robust statistics are used for the fitting of GAMLSS models (Chapter 12), bootstrapping is used for simultaneously checking the skewness and kurtosis of those models (Chapter 16; see also Wright & Herrington (2011)), and exploratory data analysis (EDA) is essentially at the core of the entire GAMLSS framework (e.g. see Chapter 17. A recent example of the value of EDA in GAMLSS can be found in Stadlmann & Kneib (2021)). The sections in the main article elaborate on the robustness of GAMLSS models to outlier contamination and the assessment of data's shape (i.e. skewness and kurtosis).

References

FRIEDRICH, S., ANTES, G., BEHR, S., BINDER, H., BRANNATH, W., DUMPERT, F., ICKSTADT, K., KESTLER, H., LEDERER, J., LEITGÖB, H., PAULY, M., STELAND, A., WILHELM, A. & FRIEDE, T. (2021). Is there a role for statistics in artificial intelligence? *Advances in Data Analysis and Classification* URL `https://doi.org/10.1007/s11634-021-00455-6`.

GELMAN, A. & VEHTARI, A. (2021). What are the most important statistical ideas of the past 50 years? *Journal of the American Statistical Association* **116**, 2087–2097. URL `10.1080/01621459.2021.1938081`.

KRISHNAMOORTHY, K. (2016). *Handbook of statistical distributions with applications.* Boca Raton: CRC Press.

MANUGUERRA, M., HELLER, G.Z. & M., J. (2020). Continuous ordinal regression for analysis of visual analogue scales: The R package ordinalCont. *Journal of Statistical Software* **96**, 1–25. doi: 10.18637/jss.v096.i08. URL `https://www.jstatsoft.org/index.php/jss/article/view/v096i08`.

MARMOLEJO-RAMOS, F., BARRERA-CAUSIL, C., KUANG, S., FAZLALI, Z., WEGENER, D., KNEIB, T., DE BASTIANI, F. & MARTINEZ-FLÓREZ, G. (in press). Generalised Exponential-Gaussian distribution: a method for neural reaction time analysis. *Cognitive Neurodynamics* .

RIGBY, R., STASINOPOULOS, M., HELLER, G. & DE BASTIANI, F. (2021). *Distributions for modeling location, scale, and shape using GAMLSS in R.* Boca Raton: CRC Press.

ROQUIM, F., RAMIRES, T., NAKAMURA, L., RIGHETTO, A., LIMA, R. & GOMES, R. (2021). Building flexible regression models: including the Birnbaum-Saunders distribution in the GAMLSS package. *Semina: Ciências Exatas e Tecnológicas* **42**, 163–168. doi:http://dx.doi.org/10.5433/1679-0375.2021v42n2p163.

STADLMANN, S. & KNEIB, T. (2021). Interactively visualizing distributional regression models with distreg.vis. *Statistical Modelling* Https://doi.org/10.1177/1471082X211007308.

STASINOPOULOS, M., RIGBY, R., HELLER, G., VOUDOURIS, V. & DE BASTIANI, F. (2017). *Flexible regression and smoothing using GAMLSS in R.* Boca Raton: CRC Press.

STASINOPOULOS, M.D., RIGBY, R.A. & DE BASTIANI, F. (2018). GAMLSS: A distributional regression approach. *Statistical Modelling* **18**, 248–273.

WELSH, A. (2019). [Review of the book "Flexible regression and smoothing using GAMLSS in R", by M. D. Stasinopoulos, R. A. Rigby, G. Z. Heller, V. Voudouris, & F. de Bastiani]. *Australian and New Zealand Journal of Statistics* **61**, 392–395.

WRIGHT, D. & HERRINGTON, J. (2011). Problematic standard errors and confidence intervals for skewness and kurtosis. *Behavior Research Methods* **43**, 8–17.