

## RESEARCH ARTICLE

# The unit ratio-extended Weibull family and the dropout rate in Brazilian undergraduate courses

Fernando A. Peña-Ramírez<sup>1\*</sup>, Renata R. Guerra<sup>2</sup>, Charles Peixoto Mafalda<sup>3</sup>

**1** Departamento de Estadística, Universidad Nacional de Colombia, Bogotá, Colombia, **2** Departamento de Estatística, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil, **3** Departamento de Estatística, Universidade Federal de Pernambuco, Recife, PE, Brazil

\* [fapenara@unal.edu.co](mailto:fapenara@unal.edu.co)



## Abstract

We propose a new family of distributions, so-called the unit ratio-extended Weibull family ( $UR\mathcal{E}\mathcal{W}$ ). It is derived from ratio transformation in an extended Weibull random variable. The use of this transformation is a novelty of the work since it has been less explored than the exponential and has not yet been studied within the extended Weibull class. Moreover, we offer a valuable alternative to model double-bounded variables on the unit interval. Five  $UR\mathcal{E}\mathcal{W}$  special models are studied in detail, namely the: i) unit ratio-Gompertz; ii) unit ratio-Burr XII; iii) unit ratio-Lomax; v) unit ratio-Rayleigh, and vi) unit ratio-Weibull distributions. We propose a quantile-parameterization for the new family. The maximum likelihood estimators (MLEs) are presented. A Monte Carlo study is performed to evaluate the behavior of the MLEs of unit ratio-Gompertz and unit ratio-Rayleigh distributions. This last model has closed-form and approximately unbiased MLE for small sample sizes. Further, the  $UR\mathcal{E}\mathcal{W}$  submodels are adjusted to the dropout rate in Brazilian undergraduate courses. We focus on the areas of civil engineering, economics, computer sciences, and control engineering. The applications show that the new family is suitable for modeling educational data and may provide effective alternatives compared to other usual unit models, such as the Beta, Kumaraswamy, and unit gamma distributions. They can also outperform some recent contributions in the unit distribution literature. Thus, the  $UR\mathcal{E}\mathcal{W}$  family can provide competitive alternatives when those models are unsuitable.

## OPEN ACCESS

**Citation:** Peña-Ramírez FA, Guerra RR, Mafalda CP (2023) The unit ratio-extended Weibull family and the dropout rate in Brazilian undergraduate courses. PLoS ONE 18(11): e0290885. <https://doi.org/10.1371/journal.pone.0290885>

**Editor:** Mohamed R. Abonazel, Cairo University, EGYPT

**Received:** April 19, 2023

**Accepted:** August 11, 2023

**Published:** November 16, 2023

**Copyright:** © 2023 Peña-Ramírez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting information files](#).

**Funding:** This research was partially funded by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Brazil, grant number 23/2551-0000851-3 awarded by RRG.

**Competing interests:** The authors have declared that no competing interests exist.

## 1 Introduction

The formulation of new generalized classes of probability distributions is a topic that has received a great deal of attention in recent years, particularly when it comes to positive data [1]. To mention a few, we refer the reader to [2–4] as extensions of the Weibull distribution and [5, 6] for Nadarajah-Haghighi generalizations. Most of these works are introduced aiming to furnish more flexible distributions regarding shape densities and hazard rates. However, there is much to be done when considering random variables supported in the unit interval. We can cite the beta and Kumaraswamy [7] ( $\mathcal{KW}$ ) distributions as classical unit models. In this respect,

Motivated by the increasing interest in modeling bounded data, other unit distributions have been introduced and are available in the literature. Some of these advances are, for instance, the unit gamma [8] ( $\mathcal{UG}$ ), simplex [9], CDF-quantile [10], unit Birnbaum-Saunders [11, 12] ( $\mathcal{UBS}$ ), unit Weibull [13] ( $\mathcal{UW}$ ), unit extended Weibull [14], complementary unit extended Weibull [14], unit Gompertz [15], unit Burr XII [16, 17], reflected unit Burr XII [18], unit generalized half normal [19], bounded odd inverse Pareto exponential [20], Modified Kumaraswamy [21], unit-sinh-normal [22], log-Bilal [23] and log-weighted exponential [24] distributions. This interest is due to several natural and anthropogenic phenomena which are bounded in a certain interval [12]. The list of double-bounded random variables may include the proportion of chemical components in different substances [25], vote proportions [12, 26], relative air humidity [27], well-being indicators [14], mortality rates [18, 28], loss given default [29] among several other indexes, indicators, ratios, and rates. Nevertheless, some situations may require other alternatives to model heavy tails and asymmetric proportion data where current models have limitations.

In this context, we introduce the so-called unit ratio-extended Weibull ( $\mathcal{UR}\mathcal{E}\mathcal{W}$ ) family of distributions, which is built upon the ratio transformation in the extended Weibull [30] ( $\mathcal{E}\mathcal{W}$ ) class. The most common method to derive those unit distributions is applying the exponential transformation in positive random variables. The use of ratio transformation is a novelty of this work since it has been less explored and has not yet been studied within the  $\mathcal{E}\mathcal{W}$  class. One advantage of introducing the  $\mathcal{UR}\mathcal{E}\mathcal{W}$  is that some special models can produce N-shaped, U-shaped, and unimodal density shapes. These features make the proposed family quite attractive for educational modeling and addressing real-life problems involving asymmetric and heavy-tailed double-bounded indicators. The N-shaped behavior, for example, is not assumed by the classical beta and  $\mathcal{KW}$  distributions but can be accommodated by some  $\mathcal{UR}\mathcal{E}\mathcal{W}$  special cases. We conduct shape analysis and provide density plots on the proposed models to illustrate these characteristics.

Our main contribution lies in offering a valuable alternative to model double-bounded variables in the unit interval. Moreover, we present at least four contributions achieved by pioneering the  $\mathcal{UR}\mathcal{E}\mathcal{W}$  class. First, the new family has more than twenty special models that may provide a source of alternatives to deal with rates and proportions, among other random variables in the unit domain. The second contribution is to provide a quantile parametrization for the new family. This framework is useful since the quantiles are outlier-resistant location measures and have a more intuitive interpretation than the original parametrization. The third contribution is related to parameter estimation under the maximum likelihood approach. As illustrated in Section 4, some one-parameter  $\mathcal{UR}\mathcal{E}\mathcal{W}$  special cases present its maximum likelihood estimator (MLE) in closed form.

Finally, the fourth contribution to formulating the  $\mathcal{UR}\mathcal{E}\mathcal{W}$  family is its applicability for modeling educational indicators. This type of data has motivated the proposal of several unit distributions. It is the case of [31–33], which examines indicators related to educational attainment percentage and school living conditions across various countries. We can also cite [14] for analyzing literacy rates in Brazilian and Colombian municipalities and [17] for modeling the dropout rate of Brazilian undergraduate animal science courses. However, it should be noted that, to the author's knowledge, there remains a significant gap in the available information regarding the phenomenon of first-year, or freshman, student dropout. This paper's motivating data sets concern the first-year dropout rate in Brazilian undergraduate courses. We analyze this outcome for civil engineering, economics, computer sciences, and control engineering courses. The four data sets are positive-skewed, and we observe a short amount of courses with a dropout rate smaller than 17%. This feature should be common for these kinds of data. When analyzing higher education institutions, academic programs with low dropout

rates tend to receive higher quality ratings [34]. In addition, this measure is seen as an indicator of institutional excellence and performance [35]. Therefore, our proposals have the advantage of providing consistently better fits than classical beta and  $\mathcal{KW}$  distributions when modeling the dropout rate in Brazilian undergraduate courses (see Section 6). As illustrated in the applications, they can also outperform some recent contributions in the distribution literature, such as the  $\mathcal{UBS}$ ,  $\mathcal{UW}$ , and  $\mathcal{CUW}$  distributions. All analysis in this paper is carried out using R programming language. The computational codes and data sets used to obtain the plots, simulations, and application results are made available on a GitHub repository (Computer codes available at <https://github.com/penaramirez/UREW>).

The rest of the paper is organized as follows. Section 2 presents the theoretical background and defines the new family of unit distributions. Some  $\mathcal{UR}\mathcal{E}\mathcal{W}$  special cases are presented in Section 3. Section 4 focuses on inferential procedures based on the maximum likelihood method. We present results for all family members and derive expressions for the MLEs of some special models. Section 5 discusses simulation studies' results to assess the performance of the point and asymptotic interval estimators. Section 6 illustrates our proposed family's relevance in educational data, specifically about the first-year dropout rate in some Brazilian undergraduate courses. The final remarks are presented in Section 7.

## 2 The unit ratio-extended Weibull family of distributions

This section presents the theoretical background and defines the proposed family from a ratio transformation in the  $\mathcal{EW}$  class of distributions. Let  $X$  be a random variable on the  $\mathcal{EW}$  class, and denote  $X \sim \mathcal{EW}(\alpha, \xi)$ . The probability density function (pdf) of  $X$  is given by

$$g(x) = \alpha h(x; \xi) \exp[-\alpha H(x; \xi)], \tag{1}$$

where  $x > 0$ ,  $\alpha > 0$ ,  $H(x; \xi)$  is a non-negative monotonically increasing function which depends on the parameter vector  $\xi$ , and  $h(x; \xi)$  is the derivative of  $H(x; \xi)$  with respect to  $x$ . For each formulation of  $H(x; \xi)$ , different  $\mathcal{EW}$  special models result. Thus, several well-known distributions can be obtained depending on the choice of this function. Table 1 presents twenty alternatives for  $H(x; \xi)$ , their corresponding derivatives, and inverse functions. Further details on this family and some generalizations to examine non-negative data are given by [36–38].

The  $\mathcal{EW}$  cumulative distribution function (cdf) and quantile function (qf) are given by

$$G(x) = 1 - \exp[-\alpha H(x; \xi)],$$

and

$$Q(u) = H^{-1}\left[-\frac{1}{\alpha} \log(1 - u)\right],$$

respectively, where  $H^{-1}(\cdot; \xi)$  is the inverse function of  $H(\cdot; \xi)$ .

We define the  $\mathcal{UR}\mathcal{E}\mathcal{W}$  class of distribution by considering the ratio transformation  $Y = X/(1 + X)$ , where  $X \sim \mathcal{EW}(\alpha, \xi)$ . Hereafter, we denote  $Y$  as a  $\mathcal{UR}\mathcal{E}\mathcal{W}$  random variable, which has cdf

$$F_Y(y) = 1 - \exp\left[-\alpha H\left(\frac{y}{1-y}; \xi\right)\right], \tag{2}$$

where  $0 < y < 1$ ,  $\alpha > 0$ , and  $\xi$  is the parameter vector associated to the  $H(\cdot; \xi)$  function.

Thus, the pdf and qf of the proposed family are

$$f_Y(y) = \alpha (1 - y)^{-2} h\left(\frac{y}{1-y}; \xi\right) \exp\left[-\alpha H\left(\frac{y}{1-y}; \xi\right)\right],$$

**Table 1. Some  $\mathcal{EW}$  special models and their corresponding  $H(x; \xi)$ ,  $H^{-1}(x; \xi)$  and  $h(x; \xi)$  functions.**

| Distribution              | $H(x; \xi)$                                       | $H^{-1}(x; \xi)$                             | $h(x; \xi)$   | $\alpha$ | $\xi$                                    |
|---------------------------|---|--|---|----------|--|
| Additive Weibull          | $(x/\beta_1)^{\alpha_1} + (x/\beta_2)^{\alpha_2}$ | no closed-form                               | $(\alpha_1/\beta_1)(x/\beta_1)^{\alpha_1-1} + (\alpha_2/\beta_2)(x/\beta_2)^{\alpha_2-1}$ | 1        | $[\alpha_1, \alpha_2, \beta_1, \beta_2]$ |
| Burr XII                  | $\log(1 - x^\beta)$                               | $[1 - \exp(x)]^{1/\beta}$                    | $\beta x^{\beta-1}/(1 + x^\beta)$   | $\alpha$ | $\beta$                                  |
| Chen                      | $\exp(x^b) - 1$                                   | $[\log(x + 1)]^{1/b}$                        | $bx^{b-1} \exp(x^b)$  | $\alpha$ | $b$                                      |
| Exponential               | $x$   | $x$  | 1   | $\alpha$ | $\emptyset$                              |
| Exponential power         | $\exp[(\lambda x)^\beta] - 1$                     | $\lambda^{-1}[\log(x + 1)]^{1/\beta}$        | $\beta \lambda \exp[(\lambda x)^\beta](\lambda x)^{\beta-1}$                              | 1        | $[\lambda, \beta]$                       |
| Flexible Weibull          | $\exp(\lambda x - \beta/x)$                       | (*)  | $\exp(\lambda x - \beta/x)(\lambda + \beta/x^2)$  | 1        | $[\lambda, \beta]$                       |
| Fréchet                   | $x^{-y}$  | $x^{-1/y}$                                   | $-yx^{-(y+1)}$  | $\alpha$ | $y$                                      |
| Gompertz ( $x > 0$ )      | $\beta^{-1}[\exp(\beta x) - 1]$                   | $\beta^{-1} \log(\beta x + 1)$               | $\exp(\beta x)$   | $\alpha$ | $\beta$                                  |
| Linear Failure rate       | $ax + bx^2/2$                                     | no closed-form                               | $a + bx$  | 1        | $[a, b]$                                 |
| Log-logistic              | $\log(1 + x^c)$                                   | $[1 - \exp(x)]^{1/c}$                        | $cx^{c-1}/(1 + x^c)$  | 1        | $c$                                      |
| Lomax                     | $\log(1 + x/\beta)$                               | $\beta[\exp(x) - 1]$                         | $1/(\beta + x)$   | $\alpha$ | $\beta$                                  |
| Modified Weibull          | $x^\lambda \exp(\lambda x)$                       | (*)  | $x^{\lambda-1} \exp(\lambda x)(\lambda + \lambda x)$                                      | $\alpha$ | $[y, \lambda]$                           |
| Pareto ( $x > k$ )        | $\log(x/k)$                                       | $k \exp(x)$                                  | $1/x$   | $\alpha$ | $k$                                      |
| Power generalized Weibull | $[1 + (x/\beta)^{\alpha_1}]^\theta - 1$           | $\beta[(x + 1)^{1/\theta} - 1]^{1/\alpha_1}$ | $(\theta \alpha_1/\beta)[1 + (x/\beta)^{\alpha_1}]^{\theta-1} (x/\beta)^{\alpha_1}$       | 1        | $[\alpha_1, \beta, \theta]$              |
| Rayleigh                  | $x^2$   | $x^{1/2}$                                    | $2x$  | $\alpha$ | $\emptyset$                              |
| Weibull                   | $x^y$   | $x^{1/y}$                                    | $yx^{y-1}$  | $\alpha$ | $y$                                      |

<https://doi.org/10.1371/journal.pone.0290885.t001>

and

$$Q_Y(u) = \frac{H^{-1}[-\alpha^{-1} \log(1 - u); \xi]}{1 + H^{-1}[-\alpha^{-1} \log(1 - u); \xi]},$$

respectively. The proposition below refers to a quantile-based parametrization for the  $\mathcal{UREW}$  family. Analogous frameworks can be found in other unit models recently introduced. See [14, 39, 40] for median-based parametrizations and [41] for a quantile-based example.

**Proposition 1.** *Let  $Y$  be a  $\mathcal{UREW}$  random variable, then its cdf can be rewritten as*

$$F_Y(y) = 1 - (1 - \tau)^{H\left(\frac{y}{1-y}; \xi\right)/H\left(\frac{q(\tau)}{1-q(\tau)}; \xi\right)}, \quad y \in (0, 1), \tag{3}$$

where  $q(\tau) \in (0, 1)$  is a location parameter which corresponds to the  $\tau$ th quantile of  $Y$ ,  $\xi$  is the parameter vector associated with  $H(\cdot; \xi)$ , and  $\tau$  is assumed as known.

*Proof.* The result in Eq (3) holds by replacing  $\alpha = \log(1 - \tau)^{-1}/H\left(\frac{q(\tau)}{1-q(\tau)}; \xi\right)$  in (2). Hence, the qf  $Y$  can be rewritten as

$$Q_Y(u) = \frac{H^{-1}\left[H\left(\frac{q(\tau)}{1-q(\tau)}; \xi\right) \frac{\log(1-u)}{\log(1-\tau)}; \xi\right]}{1 + H^{-1}\left[H\left(\frac{q(\tau)}{1-q(\tau)}; \xi\right) \frac{\log(1-u)}{\log(1-\tau)}; \xi\right]}.$$

Setting  $u = \tau$  in the above equation, we obtain that  $Q_Y(\tau) = q(\tau)$ , which concludes the proof.

Under the quantile parametrization, the  $UR\mathcal{E}\mathcal{W}$  pdf can be written as

$$f_Y(y) = \frac{\log[(1 - \tau)^{-1}]}{(1 - y)^2 H\left(\frac{q(\tau)}{1 - q(\tau)}; \xi\right)} h\left(\frac{y}{1 - y}; \xi\right) (1 - \tau)^{H\left(\frac{y}{1 - y}; \xi\right) / H\left(\frac{q(\tau)}{1 - q(\tau)}; \xi\right)}. \tag{4}$$

### 3 Some $UR\mathcal{E}\mathcal{W}$ special cases

Several well-established statistical models are special cases in the EW family. They can be considered baseline models in the  $UR\mathcal{E}\mathcal{W}$  family by replacing their corresponding  $H(\cdot; \xi)$  functions in the cdf (2). Here, we give further details on five of those models, namely: the unit ratio-Gompertz ( $UR\mathcal{G}$ ), unit ratio-Burr XII ( $UR\mathcal{B}\mathcal{X}\mathcal{I}\mathcal{I}$ ), unit ratio-Lomax ( $UR\mathcal{L}$ ), unit ratio-Weibull ( $UR\mathcal{W}$ ), and unit ratio-Rayleigh ( $UR\mathcal{R}$ ) distributions. These models are introduced using the quantile-parametrization given in Proposition 1. The  $H(\cdot; \xi)$  functions of these and several other models members of the  $UR\mathcal{E}\mathcal{W}$  family can be consulted in Table 1.

#### 3.1 The unit ratio-Gompertz distribution

The  $UR\mathcal{G}$  distribution is obtained considering the Gompertz as a baseline model in the  $UR\mathcal{E}\mathcal{W}$  family. Thus, by taking  $H(x; \xi) = \exp\{\beta x\} - 1$  in (2), the  $UR\mathcal{G}$  cdf can be written as

$$F_{UR\mathcal{G}}(y) = 1 - (1 - \tau)^{\left(\exp\left\{\frac{\beta y}{1 - y}\right\} - 1\right) / \left(\exp\left\{\frac{\beta \mu}{1 - \mu}\right\} - 1\right)},$$

where  $y \in (0, 1)$ ,  $\beta > 0$  is a shape parameter, and  $\mu \in (0, 1)$  is the  $\tau$ th quantile parameter. The corresponding pdf, qf, and hazard rate function (hrf) are

$$f_{UR\mathcal{G}}(y) = \frac{\beta \log[(1 - \tau)^{-1}] \exp\left\{\frac{\beta y}{1 - y}\right\}}{(1 - y)^2 \left[\exp\left\{\frac{\beta \mu}{1 - \mu}\right\} - 1\right]} \times (1 - \tau)^{\left(\exp\left\{\frac{\beta y}{1 - y}\right\} - 1\right) / \left(\exp\left\{\frac{\beta \mu}{1 - \mu}\right\} - 1\right)}, \tag{5}$$

$$Q_{UR\mathcal{G}}(u) = \frac{\log\left\{\frac{\log(1 - u)}{\log(1 - \tau)} \left[\exp\left(\frac{\beta \mu}{1 - \mu}\right) - 1\right] + 1\right\}}{\beta + \log\left\{\frac{\log(1 - u)}{\log(1 - \tau)} \left[\exp\left(\frac{\beta \mu}{1 - \mu}\right) - 1\right] + 1\right\}}, \tag{6}$$

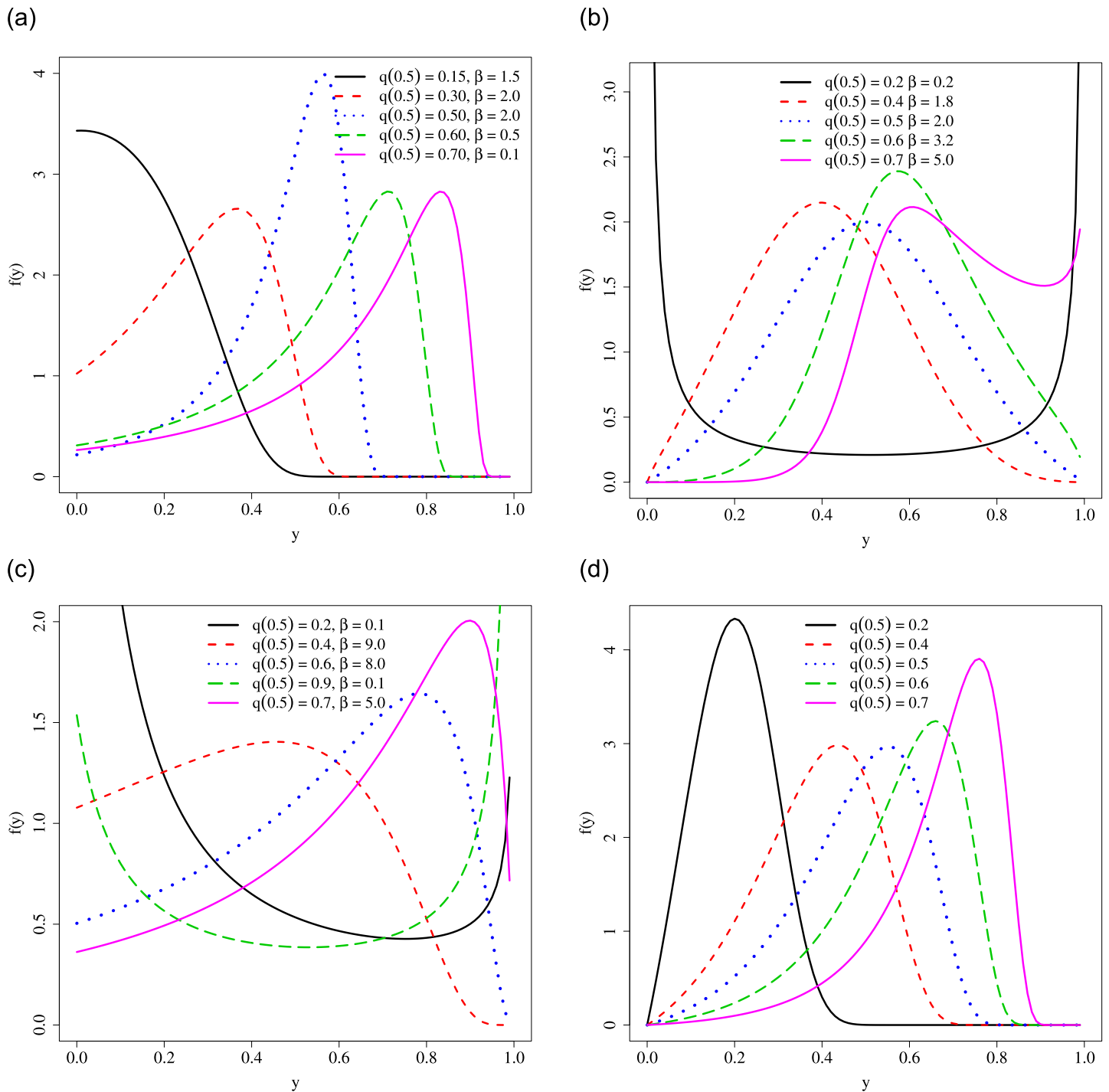
and

$$h_{UR\mathcal{G}}(y) = \frac{\beta \log[(1 - \tau)^{-1}] \exp\left\{\frac{\beta y}{1 - y}\right\}}{(1 - y)^2 \left[\exp\left\{\frac{\beta \mu}{1 - \mu}\right\} - 1\right]},$$

respectively. Fig 1(a) illustrates the  $UR\mathcal{G}$  pdf shapes for several combinations of  $\mu$  and  $\beta$ , with  $\tau = 0.5$ .

#### 3.2 The unit ratio-Burr XII distribution

The  $UR\mathcal{B}\mathcal{X}\mathcal{I}\mathcal{I}$  distribution is obtained considering the Burr XII as a baseline model in the  $UR\mathcal{E}\mathcal{W}$  family. Thus, by taking  $H(x; \xi) = \log[1 + x^\beta]$  in (2), and after simplification, the



**Fig 1. Density plots for some UREW special models. (a) URG, (b) URBXII, (c) URL, (d) URR.**

<https://doi.org/10.1371/journal.pone.0290885.g001>

URBXII cdf reduces to

$$F_{URBXII}(y) = 1 - \left[ 1 + \frac{y^\beta}{(1-y)^\beta} \right]^{\log(1-\tau)/\log\{1+q(\tau)^\beta[1-q(\tau)]^{-\beta}\}}$$

where  $y \in (0, 1)$ , and  $\mu \in (0, 1)$  is the  $\tau$ th quantile parameter. The corresponding pdf, qf, and hrf are

$$f_{URBXII}(y) = \frac{\beta y^{\beta-1} \log[(1-\tau)^{-1}]}{(1-y)^{\beta+1} \log\{1+q(\tau)^\beta [1-q(\tau)]^{-\beta}\}} \times \left[1 + \frac{y^\beta}{(1-y)^\beta}\right]^{\log(1-\tau)/\log\{1+q(\tau)^\beta [1-q(\tau)]^{-\beta}\} - 1}, \tag{7}$$

$$Q_{URBXII}(u) = \frac{\left[(1-u)^{\log\{1+q(\tau)^\beta [1-q(\tau)]^{-\beta}\} \log(1-\tau)} - 1\right]^{1/\beta}}{1 + \left[(1-u)^{\log\{1+q(\tau)^\beta [1-q(\tau)]^{-\beta}\} \log(1-\tau)} - 1\right]^{1/\beta}}, \tag{8}$$

and

$$h_{URBXII}(y) = \frac{\beta y^{\beta-1} \log[(1-\tau)^{-1}]}{(1-y)^{\beta+1} \log\{1+q(\tau)^\beta [1-q(\tau)]^{-\beta}\}},$$

respectively. Fig 1(b) illustrates the  $URBXII$  pdf shapes for several combinations of  $\mu$  and  $\tau = 0.5$ . This plot illustrates the flexibility of the  $URBXII$  distribution. It can have N-shaped, U-shaped, and unimodal density shapes, being able to fit asymmetric and heavy-tailed double-bounded data.

### 3.3 The unit ratio-Lomax distribution

The  $URL$  distribution is obtained considering the Lomax as a baseline model in the  $UREW$  family. Thus, by taking  $H(x; \xi) = \log[1+x^\beta]$  in (2), and after simplification, the  $URL$  cdf reduces to

$$F_{URL}(y) = 1 - \left[1 + \frac{y}{\beta(1-y)}\right]^{\log(1-\tau)/\log\left[1+\frac{\mu}{\beta(1-\mu)}\right]}$$

where  $y \in (0, 1)$ ,  $\beta > 0$  is a shape parameter, and  $\mu \in (0, 1)$  is the  $\tau$ th quantile parameter. The corresponding pdf, qf, and hrf are

$$f_{URL}(y) = \frac{\log[(1-\tau)^{-1}]}{(1-y)[\beta(1-y) + y] \log\left[1 + \frac{\mu}{\beta(1-\mu)}\right]} \times \left[1 + \frac{y}{\beta(1-y)}\right]^{\log(1-\tau)/\log\left[1+\frac{\mu}{\beta(1-\mu)}\right]}, \tag{9}$$

$$Q_{URL}(u) = \frac{\beta(1-u)^{\log\left[1+\frac{\mu}{\beta(1-\mu)}\right]/\log(1-\tau)} - \beta}{1 + \beta(1-u)^{\log\left[1+\frac{\mu}{\beta(1-\mu)}\right]/\log(1-\tau)} - \beta},$$

and

$$h_{URL}(y) = \frac{\log[(1-\tau)^{-1}]}{(1-y)[\beta(1-y) + y] \log\left[1 + \frac{\mu}{\beta(1-\mu)}\right]},$$

respectively. Fig 1(c) illustrates the  $UR\mathcal{L}$  pdf shapes for several combinations of  $\mu$  and  $\beta$ , with  $\tau = 0.5$ .

### 3.4 The unit ratio-Weibull and unit ratio-Rayleigh distributions

The  $UR\mathcal{W}$  distribution is obtained considering the Weibull as baseline model in the  $UR\mathcal{E}\mathcal{W}$  family. By taking  $H(x; \xi) = x^\beta$  in (2), the  $UR\mathcal{W}$  can be written as

$$F_{UR\mathcal{W}}(y) = 1 - (1 - \tau)^{y^\beta (1-\mu)^\beta / [\mu^\beta (1-y)^\beta]},$$

where  $y \in (0, 1)$ ,  $\beta > 0$  is a shape parameter, and  $\mu \in (0, 1)$  is the  $\tau$ th quantile parameter. The corresponding pdf, qf, and hrf are

$$f_{UR\mathcal{W}}(y) = \frac{\beta y^{\beta-1} (1 - \mu)^\beta}{\mu^\beta (1 - y)^{\beta+1}} [\log(1 - \tau)^{-1}] (1 - \tau)^{y^\beta (1-\mu)^\beta / [\mu^\beta (1-y)^\beta]}, \tag{10}$$

$$Q_{UR\mathcal{W}}(u) = \frac{\left[ \frac{\mu^\beta \log(1-u)}{(1-\mu)^\beta \log(1-\tau)} \right]^{1/\beta}}{1 + \left[ \frac{\mu^\beta \log(1-u)}{(1-\mu)^\beta \log(1-\tau)} \right]^{1/\beta}},$$

and

$$h_{UR\mathcal{W}}(y) = \frac{\beta y^{\beta-1} (1 - \mu)^\beta}{\mu^\beta (1 - y)^{\beta+1}} [\log(1 - \tau)^{-1}],$$

respectively. For  $\beta = 2$ , the  $UR\mathcal{W}$  reduces to the  $UR\mathcal{R}$  distribution, which is also new. The  $UR\mathcal{R}$  is a one-parameter model obtained considering the Rayleigh as a baseline model in the  $UR\mathcal{E}\mathcal{W}$  family. Its pdf is given by

$$f_{UR\mathcal{R}}(y) = \frac{2 y (1 - \mu)^2}{\mu^2 (1 - y)^3} \log(1 - \tau)^{-1} (1 - \tau)^{y^2 (1-\mu)^2 / [\mu^2 (1-y)^2]}. \tag{11}$$

Fig 1(d) illustrates the  $UR\mathcal{R}$  pdf shapes for several combinations of  $\mu$  and  $\tau = 0.5$ . It shows that the  $UR\mathcal{R}$  distribution presents unimodal density shape, accomodating left and right-skewed data in the unit interval.

## 4 Maximum likelihood estimation

Here, we consider estimation of the parameters of the  $UR\mathcal{E}\mathcal{W}$  family by the maximum likelihood (ML) method. The log-likelihood for a random sample  $y_1, \dots, y_n$  from (4), based on parameter vector  $\theta = (\mu, \xi^\top)^\top$ , is

$$\begin{aligned} \ell(\theta|y_1, \dots, y_n) &= n \log[\log(1 - \tau)^{-1}] - n \log \left[ H\left(\frac{\mu}{1-\mu}; \xi\right) \right] \\ &- 2 \sum_{i=1}^n \log(1 - y_i) + \sum_{i=1}^n \log \left[ h\left(\frac{y_i}{1-y_i}; \xi\right) \right] - \frac{\log(1 - \tau)^{-1}}{H\left(\frac{\mu}{1-\mu}; \xi\right)} \sum_{i=1}^n H\left(\frac{y_i}{1-y_i}; \xi\right). \end{aligned} \tag{12}$$

The components of the score vector  $U(\theta) = [U_\mu, U_\xi]^\top$ , are

$$U_\mu = \frac{h\left(\frac{\mu}{1-\mu}; \xi\right)}{(1 - \mu)^2 H\left(\frac{\mu}{1-\mu}; \xi\right)} \left[ -n + \frac{\log(1 - \tau)^{-1}}{H\left(\frac{\mu}{1-\mu}; \xi\right)} \sum_{i=1}^n H\left(\frac{y_i}{1-y_i}; \xi\right) \right]$$



e

$$U_{\xi} = \frac{\partial \ell}{\partial \xi} = \sum_{i=1}^n \frac{1}{h\left(\frac{y_i}{1-y_i}; \xi\right)} \frac{\partial \ell}{\partial \xi} h\left(\frac{y_i}{1-y_i}; \xi\right) - n \frac{h\left(\frac{\mu}{1-\mu}; \xi\right)}{H\left(\frac{\mu}{1-\mu}; \xi\right)} - \frac{\log(1-\tau)^{-1}}{H\left(\frac{\mu}{1-\mu}; \xi\right)} \\ \times \sum_{i=1}^n \frac{\partial \ell}{\partial \xi} h\left(\frac{y_i}{1-y_i}; \xi\right) + \frac{\log(1-\tau)^{-1} h\left(\frac{\mu}{1-\mu}; \xi\right)}{\left[H\left(\frac{\mu}{1-\mu}; \xi\right)\right]^2} \sum_{i=1}^n \frac{\partial \ell}{\partial \xi} H\left(\frac{y_i}{1-y_i}; \xi\right).$$

For fixed values of  $\xi$ , it is possible to obtain a closed-form for the MLE of the  $\mu$ . By setting  $U_{\mu} = 0$  and solving for  $\mu$ , we have

$$\hat{\mu}(\hat{\xi}) = \frac{H^{-1}\left[\frac{1}{n} \log(1-\tau)^{-1} \sum_{i=1}^n \frac{y_i}{1-y_i}; \hat{\xi}\right]}{1 + H^{-1}\left[\frac{1}{n} \log(1-\tau)^{-1} \sum_{i=1}^n \frac{y_i}{1-y_i}; \hat{\xi}\right]}.$$

Therefore, obtaining the EMV of  $\mu$  in closed-form is possible when  $\xi = \emptyset$ . Otherwise, to get the MLEs of the parameters  $\mu$  and  $\xi$ , it is necessary to use some iterative procedures such as Newton-Raphson type algorithms to maximize (12).

We can construct approximate confidence intervals for  $\theta$  based on the asymptotic normality property. Under standard regularity conditions, the asymptotic distribution of  $\hat{\theta} - \theta$  can be approximated by the multivariate normal  $\mathcal{N}(\mathbf{0}, \mathbf{J}(\hat{\theta})^{-1})$  distribution, where  $\mathbf{J}(\hat{\theta})^{-1} = -\partial \ell(\theta) / \partial \theta \partial \theta^T |_{\theta=\hat{\theta}}$  is the observed information matrix. Thus, the asymptotic 100(1 -  $\eta$ )% confidence intervals of  $\theta$  are given by

$$\hat{\theta} \pm z_{\eta/2} \times [\widehat{\text{var}}(\hat{\theta})]^{1/2},$$

where  $z_{\eta/2}$  is the quantile  $\eta/2$  of the standard normal distribution, and  $\widehat{\text{var}}(\hat{\theta}) = \text{diag}\{\mathbf{J}(\hat{\theta})^{-1}\}$ . In what follows, we present the likelihood estimation of some special cases of the  $\mathcal{UR}\mathcal{E}\mathcal{W}$  family.

### 4.1 MLE for the $\mathcal{UR}\mathcal{R}(\mu)$ distribution

Let  $y_1, \dots, y_n$  be a random sample of size  $n$  from the  $\mathcal{UR}\mathcal{R}(\mu)$  distribution. The log-likelihood function is

$$\ell(\mu|y_1, \dots, y_n) = n \log(2) + \sum_{i=1}^n \log(y_i) + n \log[(1-\mu)^2] + n \log[\log(1-\tau)^{-1}] \\ - 2n \log(\mu) - 3 \sum_{i=1}^n \log(1-y_i) - \frac{\log(1-\tau)^{-1}(1-\mu)^2}{\mu^2} \sum_{i=1}^n \frac{y_i^2}{(1-y_i)^2}.$$

The score function  $U_{\mu}$  is

$$U_{\mu} = \frac{\partial \ell}{\partial \mu} = \frac{2}{\mu(1-\mu)} \left[ -n + \frac{(1-\mu)^2 \log(1-\tau)^{-1}}{\mu^2} \sum_{i=1}^n \frac{y_i^2}{(1-y_i)^2} \right]. \tag{13}$$

By setting  $U_\mu = 0$  and solving for  $\mu$ , we have the EMV of  $\mu$  as

$$\hat{\mu} = \frac{\sqrt{\log(1 - \tau)^{-1} \sum_{i=1}^n \frac{y_i^2}{(1-y_i)^2}}}{\sqrt{n} + \sqrt{\log(1 - \tau)^{-1} \sum_{i=1}^n \frac{y_i^2}{(1-y_i)^2}}}, \tag{14}$$

and the Fisher’s observed information is computed as

$$I(\mu) = \frac{\partial^2 \ell}{\partial \mu^2} = \frac{2}{\mu^4} \left[ \frac{n\mu^2(1 - 2\mu)}{(1 - \mu)^2} + (2\mu - 3) \log(1 - \tau)^{-1} \sum_{i=1}^n \frac{y_i^2}{(1 - y_i)^2} \right]. \tag{15}$$

The conditions for the maximum value of the function  $\ell(\mu|y_1, \dots, y_n)$  require that  $I(\hat{\mu}) < 0$ . This is easily observed by substituting (14) into (15), where it is verified that

$$I(\hat{\mu}) = - \frac{4 \left( \sqrt{n} + \sqrt{\log(1 - \tau)^{-1} \sum_{i=1}^n \frac{y_i^2}{(1-y_i)^2}} \right)^4}{\log(1 - \tau)^{-1} \sum_{i=1}^n \frac{y_i^2}{(1-y_i)^2}} < 0.$$

### 4.2 MLE for the $\mathcal{URG}(\beta, \mu)$ distribution

Let  $y_1, \dots, y_n$  be a random sample of size  $n$  from the  $\mathcal{URG}$  distribution with parameter vector  $\theta = (\beta, \mu)^\top$ . The log-likelihood function is

$$\begin{aligned} \ell(\theta|y_1, \dots, y_n) &= n \log[\log(1 - \tau)^{-1}] - n \log(\beta) - \frac{n\beta\mu}{1 - \mu} + \frac{\log[\log(1 - \tau)^{-1}]}{\exp\left\{\frac{\beta\mu}{1 - \mu}\right\} - 1} \\ &\times \left[ \sum_{i=1}^n \exp\left\{\frac{\beta y_i}{1 - y_i}\right\} + n \right] - \sum_{i=1}^n \log \left[ (1 - y_i)^2 \left( \exp\left\{\frac{\beta\mu}{1 - \mu}\right\} - 1 \right) \right]. \end{aligned} \tag{16}$$

The components of the score vector  $U_\theta = (U_\beta, U_\mu)^\top$  are

$$\begin{aligned} U_\beta &= \frac{\partial \ell}{\partial \beta} = \frac{n\mu}{1 - \mu} - n \frac{\exp\left\{\frac{\beta\mu}{1 - \mu}\right\} - 1}{\beta} + \frac{\mu}{1 - \mu} \log(1 - \tau)^{-1} \sum_{i=1}^n \exp\left\{\frac{\beta y_i}{1 - y_i}\right\} \\ &+ \log(1 - \tau)^{-1} \sum_{i=1}^n \frac{y_i}{1 - y_i} \exp\left\{\frac{\beta y_i}{1 - y_i}\right\}, \end{aligned}$$

and

$$\begin{aligned} U_\mu &= \frac{\partial \ell}{\partial \mu} = \frac{\log[\log(1 - \tau)^{-1}]}{\left[\exp\left\{\frac{\beta\mu}{1 - \mu}\right\} - 1\right]^2} \sum_{i=1}^n \exp\left\{\frac{\beta y_i}{1 - y_i}\right\} - n - \frac{\exp\left\{\frac{\beta\mu}{1 - \mu}\right\}}{\exp\left\{\frac{\beta\mu}{1 - \mu}\right\} - 1} \\ &- n \log[\log(1 - \tau)^{-1}]. \end{aligned}$$

Note that the system of equations  $U_\theta = \mathbf{0}$  cannot be solved in closed form; therefore, the maximization of (16) to obtain the EMV of  $\theta$  can be carried out using the quasi-Newton BFGS non-linear optimization algorithm implemented in the `optim` function available in R.

## 5 Simulation study

In this Section, a Monte Carlo study is carried on to evaluate the performance of the MLEs of the  $UREW$  family in finite samples. For that, the  $URR$  and  $URG$  distributions are considered. This study conducted 10,000 Monte Carlo replications with sample sizes  $n \in \{10, 25, 50, 75, 100\}$ . Aiming to evaluate the point estimators, we use the set of estimates of parameters obtained in each replication to calculate its mean, variance, relative biases (RBs), standard deviations (SDs), and root mean squared errors (RMSEs). Regarding the initial values selected for simulation, we highlight that the  $URR$  distribution has a closed form for its MLE (see Eq (14)). Therefore, one advantage of using this model is that it does not require defining initial values in the ML method. For the two-parameter  $UREW$  special cases, we use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm and compute the observed information matrix numerically from the optim function in the R programming language. Therefore, we set the sample quantile as the initial value for  $\mu$  and one for the shape parameter. These values are used either for the simulated or actual data experiments performed in the paper. We calculate the coverage probability of the 95% pointwise confidence interval ( $CP_{95\%}$ ) to evaluate the interval estimation. Next, we provide the numerical results for both considered distributions. Next, the numerical results for both distributions considered are presented.

### 5.1 Numerical Analysis for the $URR$ distribution

We generate occurrences of the variable  $Y$  following a  $URR$  law with five different values of  $\mu$  (scenarios). For that, we use the inversion method replacing  $u \sim U(0, 1)$  in the  $URR$  qf. The simulation results are shown in Table 2. It reveals low RB values in all the scenarios and sample sizes considered. We highlight that all its observed values are less than 0.7%. We also observe low SD values, all less than 0.5. For all the sample sizes, it is common to observe RMSE's lower values for the central values of  $\mu$  ( $\mu = 0.4$ , for example) than for the close values of the extremes ( $\mu = 0.15$  or  $\mu = 0.9$ , for example). In its last column, it can verify that the coverage probabilities of the 95% pointwise confidence intervals of the parameter are quite close to the nominal level.

Fig 2 indicates that the RB and RMSE of  $\hat{\mu}$  decrease as the sample size increases, corroborating the asymptotic properties of the MLEs.

### 5.2 Numerical analysis for the $URG$ distribution

Analogous to the previous experiment, occurrences of the variable  $Y$  are initially generated, which follows a  $URG$  distribution with different configurations in its parameters  $\mu$  and  $\beta$ . The data are generated using the inversion method in the  $URG$  qf. In Table 3, we present the simulation results. It shows that  $\mu$ 's estimates are more accurate than  $\beta$ 's. We can also observe that the RB of  $\mu$  is always less than 0.4% in absolute value. For sample sizes greater than 75, the RB of  $\hat{\beta}$  is always less than 10%. In the last column of Table 2, we can be observed that the coverage probabilities of the 95% pointwise confidence intervals of both parameters are quite close to the nominal level.

Fig 3(a) presents a plot with the sum of the RB of  $\hat{\mu}$  and  $\hat{\beta}$  that we call the total RB. Fig 3(b) presents a similar plot with the sum of the RMSE of  $\hat{\mu}$  and  $\hat{\beta}$ , that we call the total RMSE. They show that the total RB and total RMSE of  $\hat{\mu}$  and  $\hat{\beta}$  decrease as the sample size increases, corroborating the asymptotic properties of the MLEs.

## 6 Applications

This section illustrates the usefulness of the  $UREW$  family through applications in educational data related to student dropout, also known as student attrition. This outcome has some

Table 2. Results of the Monte Carlo simulation from the  $UR\mathcal{R}$  distribution.

| Scenario | $\mu$ | $n$ | mean   | Variance | RB %    | RMSE   | $CP_{95\%}$ |
|----------|-------|-----|--------|----------|---------|--------|-------------|
| 1        | 0.15  | 10  | 0.1475 | 0.0004   | -0.2432 | 0.0003 | 0.9490      |
|          |       | 25  | 0.1492 | 0.0001   | -0.0741 | 0.0001 | 0.9430      |
|          |       | 50  | 0.1493 | 0.0001   | -0.0617 | 0.0001 | 0.9590      |
|          |       | 75  | 0.1494 | 0.0001   | -0.0505 | 0.0001 | 0.9480      |
|          |       | 100 | 0.1497 | 0.0001   | -0.0286 | 0.0001 | 0.9580      |
| 2        | 0.25  | 10  | 0.2463 | 0.0008   | -0.3669 | 0.0009 | 0.9470      |
|          |       | 25  | 0.2484 | 0.0003   | -0.1584 | 0.0003 | 0.9579      |
|          |       | 50  | 0.2492 | 0.0001   | -0.0773 | 0.0001 | 0.9533      |
|          |       | 75  | 0.2495 | 0.0001   | -0.0446 | 0.0001 | 0.9537      |
|          |       | 100 | 0.2497 | 0.0001   | -0.0287 | 0.0001 | 0.9518      |
| 3        | 0.4   | 10  | 0.3943 | 0.0014   | -0.5635 | 0.0015 | 0.9552      |
|          |       | 25  | 0.3976 | 0.0005   | -0.2385 | 0.0005 | 0.9594      |
|          |       | 50  | 0.3988 | 0.0002   | -0.1170 | 0.0002 | 0.9538      |
|          |       | 75  | 0.3993 | 0.0002   | -0.0689 | 0.0001 | 0.9536      |
|          |       | 100 | 0.3995 | 0.0001   | -0.0459 | 0.0001 | 0.9508      |
| 4        | 0.7   | 10  | 0.6933 | 0.0012   | -0.6654 | 0.0012 | 0.9518      |
|          |       | 25  | 0.6972 | 0.0004   | -0.2725 | 0.0004 | 0.9535      |
|          |       | 50  | 0.6986 | 0.0002   | -0.1341 | 0.0002 | 0.9506      |
|          |       | 75  | 0.6991 | 0.0001   | -0.0812 | 0.0001 | 0.9521      |
|          |       | 100 | 0.6994 | 0.0001   | -0.0562 | 0.0001 | 0.9469      |
| 5        | 0.9   | 10  | 0.8966 | 0.0002   | -0.3384 | 0.0002 | 0.9405      |
|          |       | 25  | 0.8986 | 0.0001   | -0.1356 | 0.0001 | 0.9489      |
|          |       | 50  | 0.8993 | 0.0001   | -0.0666 | 0.0001 | 0.9497      |
|          |       | 75  | 0.8995 | 0.0001   | -0.0408 | 0.0001 | 0.9515      |
|          |       | 100 | 0.8997 | 0.0001   | -0.0287 | 0.0001 | 0.9454      |

<https://doi.org/10.1371/journal.pone.0290885.t002>

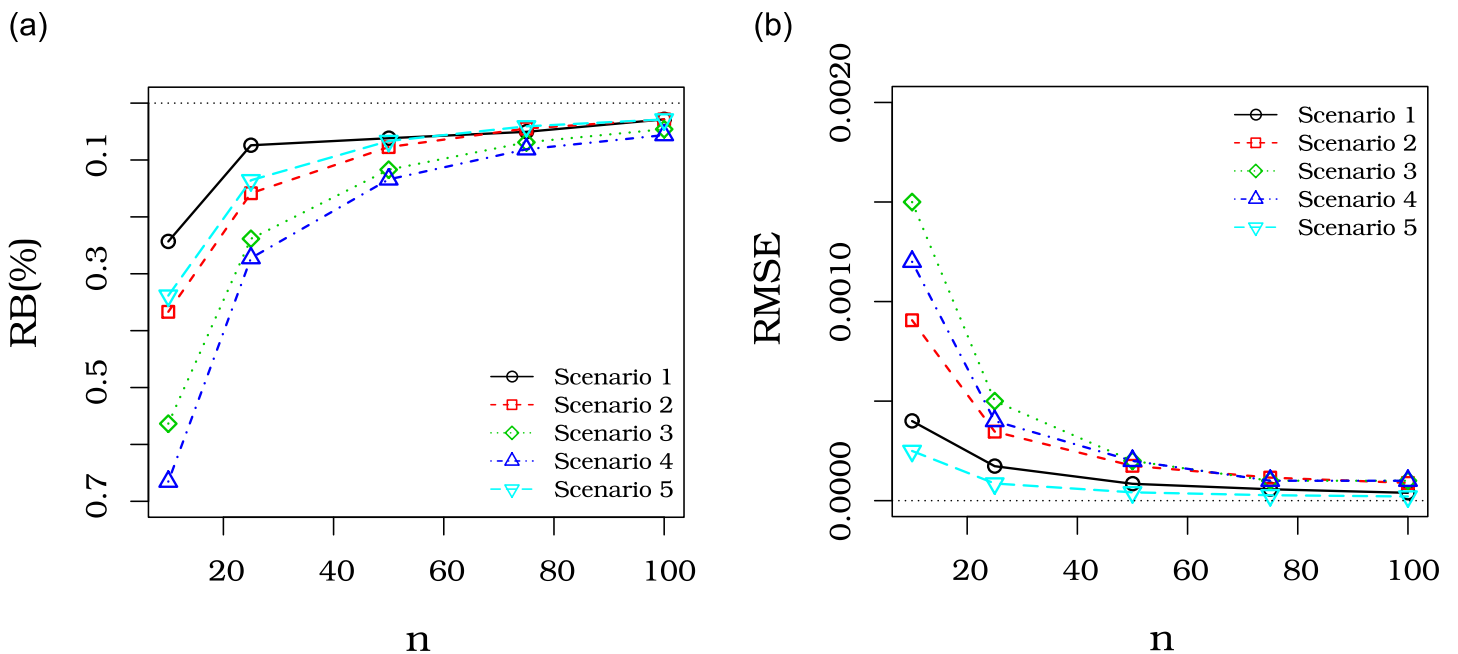


Fig 2. Percentual RB and RMSE of the  $UR\mathcal{R}$  estimator in several scenarios.

<https://doi.org/10.1371/journal.pone.0290885.g002>

Table 3. Results of the Monte Carlo simulation from the  $UR\mathcal{G}$  distribution.

| Scenario | $\mu$ | $\beta$ | $n$ | Mean        |               | Variance    |               | RB %( $\hat{\mu}$ ) |               | RMSE( $\hat{\mu}$ ) |               | CP <sub>95%</sub> |               |
|----------|-------|---------|-----|-------------|---------------|-------------|---------------|---------------------|---------------|---------------------|---------------|-------------------|---------------|
|          |       |         |     | $\hat{\mu}$ | $\hat{\beta}$ | $\hat{\mu}$ | $\hat{\beta}$ | $\hat{\mu}$         | $\hat{\beta}$ | $\hat{\mu}$         | $\hat{\beta}$ | $\hat{\mu}$       | $\hat{\beta}$ |
| 1        | 0.4   | 0.8     | 10  | 0.4038      | 1.2889        | 0.0046      | 0.7510        | 0.3819              | 48.8994       | 0.0047              | 0.9902        | 0.8852            | 0.9311        |
|          |       |         | 25  | 0.4014      | 0.9674        | 0.0019      | 0.1895        | 0.1444              | 16.7461       | 0.0019              | 0.2176        | 0.9268            | 0.9433        |
|          |       |         | 50  | 0.4011      | 0.8856        | 0.0010      | 0.0827        | 0.1146              | 8.5658        | 0.0010              | 0.0901        | 0.9377            | 0.9474        |
|          |       |         | 75  | 0.4009      | 0.8571        | 0.0006      | 0.0526        | 0.0995              | 5.7184        | 0.0006              | 0.0559        | 0.9410            | 0.9489        |
|          |       |         | 100 | 0.4006      | 0.8400        | 0.0005      | 0.0384        | 0.0653              | 4.0097        | 0.0005              | 0.0400        | 0.9423            | 0.9462        |
| 2        | 0.5   | 1.5     | 10  | 0.4971      | 1.9084        | 0.0025      | 0.6566        | -0.2869             | 40.8414       | 0.0025              | 0.8234        | 0.8941            | 0.9427        |
|          |       |         | 25  | 0.4985      | 1.6408        | 0.0010      | 0.1832        | -0.1444             | 14.0812       | 0.0010              | 0.2031        | 0.9283            | 0.9480        |
|          |       |         | 50  | 0.4994      | 1.5709        | 0.0004      | 0.0824        | -0.0544             | 7.0981        | 0.0004              | 0.0874        | 0.9406            | 0.9499        |
|          |       |         | 75  | 0.4998      | 1.5485        | 0.0003      | 0.0534        | -0.0195             | 4.8538        | 0.0003              | 0.0558        | 0.9412            | 0.9498        |
|          |       |         | 100 | 0.4997      | 1.5333        | 0.0002      | 0.0393        | -0.0226             | 3.3385        | 0.0002              | 0.0404        | 0.9444            | 0.9476        |
| 3        | 0.6   | 2.8     | 10  | 0.5980      | 3.2892        | 0.0004      | 1.0709        | -0.1989             | 48.9285       | 0.0004              | 1.3103        | 0.9042            | 0.9526        |
|          |       |         | 25  | 0.5991      | 2.9682        | 0.0001      | 0.2841        | -0.0875             | 16.8275       | 0.0001              | 0.3124        | 0.9339            | 0.9508        |
|          |       |         | 50  | 0.5996      | 2.8829        | 0.0008      | 0.1265        | -0.0388             | 8.2924        | 0.0008              | 0.1333        | 0.9430            | 0.9513        |
|          |       |         | 75  | 0.5998      | 2.8571        | 0.0001      | 0.0816        | -0.0182             | 5.7180        | 0.0001              | 0.0849        | 0.9437            | 0.9510        |
|          |       |         | 100 | 0.5998      | 2.8390        | 0.0001      | 0.0601        | -0.0167             | 3.9059        | 0.0001              | 0.0616        | 0.9456            | 0.9475        |
| 4        | 0.7   | 5       | 10  | 0.6996      | 5.8574        | 0.0001      | 3.0886        | -0.0362             | 85.7439       | 0.0001              | 3.8238        | 0.9016            | 0.9499        |
|          |       |         | 25  | 0.6998      | 5.2956        | 0.0001      | 0.7867        | -0.0164             | 29.5603       | 0.0001              | 0.8741        | 0.9350            | 0.9507        |
|          |       |         | 50  | 0.6999      | 5.1454        | 0.0001      | 0.3466        | -0.0066             | 14.5412       | 0.0001              | 0.3677        | 0.9436            | 0.9504        |
|          |       |         | 75  | 0.6999      | 5.0996        | 0.0001      | 0.2220        | -0.0018             | 9.9651        | 0.0001              | 0.2319        | 0.9435            | 0.9505        |
|          |       |         | 100 | 0.6999      | 5.0688        | 0.0001      | 0.1630        | -0.0022             | 6.8801        | 0.0001              | 0.1677        | 0.9449            | 0.9481        |
| 5        | 0.7   | 0.9     | 10  | 0.6961      | 1.1001        | 0.0012      | 0.1686        | -0.3809             | 20.0002       | 0.0012              | 0.2086        | 0.8990            | 0.9485        |
|          |       |         | 25  | 0.6983      | 0.9690        | 0.0004      | 0.0471        | -0.1693             | 6.9054        | 0.0004              | 0.0519        | 0.9297            | 0.9507        |
|          |       |         | 50  | 0.6992      | 0.9345        | 0.0002      | 0.0212        | -0.0738             | 3.4534        | 0.0002              | 0.0224        | 0.9420            | 0.9513        |
|          |       |         | 75  | 0.6996      | 0.9238        | 0.0001      | 0.0138        | -0.0374             | 2.3819        | 0.0001              | 0.0143        | 0.9415            | 0.9507        |
|          |       |         | 100 | 0.6996      | 0.9162        | 0.0001      | 0.0101        | -0.0337             | 1.6285        | 0.0001              | 0.0104        | 0.9446            | 0.9486        |
| 6        | 0.75  | 1.1     | 10  | 0.7475      | 1.3025        | 0.0004      | 0.1828        | -0.2473             | 20.2578       | 0.0004              | 0.2239        | 0.9037            | 0.9520        |
|          |       |         | 25  | 0.7489      | 1.1697        | 0.0001      | 0.0496        | -0.1068             | 6.9781        | 0.0001              | 0.0545        | 0.9330            | 0.9515        |
|          |       |         | 50  | 0.7495      | 1.1345        | 0.0001      | 0.0222        | -0.0477             | 3.4546        | 0.0001              | 0.0234        | 0.9432            | 0.9516        |
|          |       |         | 75  | 0.7497      | 1.1238        | 0.0001      | 0.0144        | -0.0244             | 2.3864        | 0.0001              | 0.0149        | 0.9429            | 0.9515        |
|          |       |         | 100 | 0.7497      | 1.1163        | 0.0001      | 0.0106        | -0.0214             | 1.6317        | 0.0001              | 0.0108        | 0.9452            | 0.9470        |

<https://doi.org/10.1371/journal.pone.0290885.t003>

complexity in data collection [42], and a diversity of definitions has been considered in the specialized literature. In this paper, we are interested in analyzing the first-year dropout rate in undergraduate courses, defined as the proportion of students who withdraw from the course before completing the first year. Thus, from a sample with  $n$  undergraduate courses, the  $i$ th observation is obtained as

$$\text{DROPOUT\_RATE}_i = \frac{\text{number of freshmen students who dropped out the } i\text{th course before completing the first year}}{\text{number of freshmen students enrolled in the } i\text{th course}},$$

where  $i \in \{1, \dots, n\}$ . The decision to focus the study on freshmen students lies in the evidence that the risk of dropping out is higher during the first year of college, also called the freshmen year [42, 43]. This period is seen as the most critical time for the connection between academic programs and students [44]. Therefore, understanding the behavior of this variable may be

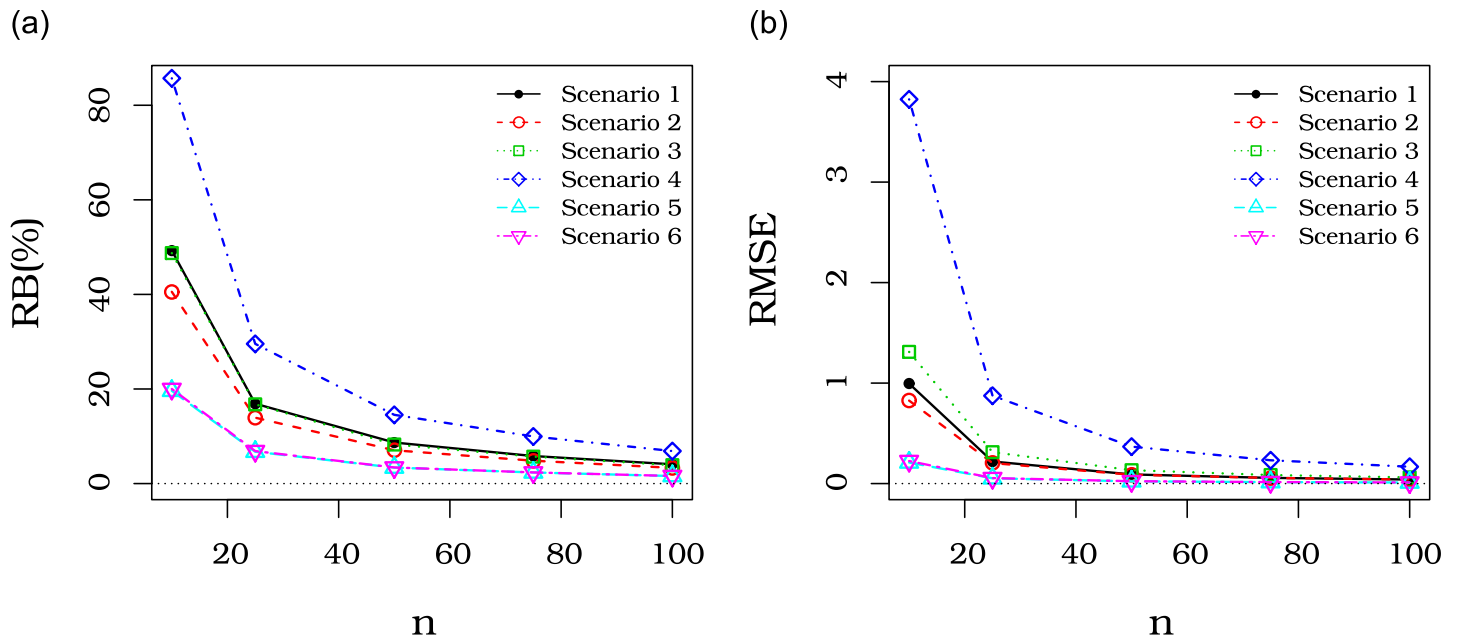


Fig 3. Total percentual RB and total RMSE of the  $UR\mathcal{G}$  estimators in several scenarios. (a) VR(%) total, (b) EQM total.

<https://doi.org/10.1371/journal.pone.0290885.g003>

helpful in developing practices aimed at reducing the early dropout from undergraduate courses from different areas.

The data used in this case study were collected from the Brazilian higher education census microdata, conducted in 2018 [45] and were calculated from the entering students in 2018. We select the presential courses with more than 29 new students and first-year dropout rate in the (0, 1) interval in the census academic year. The applications refer to four data sets about civil engineering, economics, computer sciences, and control engineering courses. We fit  $UR\mathcal{E}W$  special models and compare their performance with other existing double-bounded distributions, which are not special cases of the proposed family.

Table 4 gives a descriptive summary of the dropout rates of each dataset considered. The Economics course exhibits smaller values for all central tendency measures and higher for the skewness, kurtosis, and amplitude measures. The other courses present those measures quite close when compared with each other. Their mean and median are around 17% and 16%, respectively. The descriptive measures indicate that, for all data sets, the mass of observations concentrates on the left. This configuration is adequate since the dropout rate is negatively related to institutional quality and effectiveness. Academic programs with low dropout rates are often considered to be more efficient [34]. Nevertheless, the dropout rates in higher education are social and institutional concerns [42], and there is a broad consensus on the need for

Table 4. Descriptive statistics for the dropout rates in the four course types considered.

| Course type         | Mean   | Median | Variance | Skewness | Kurtosis | Min.   | Max.   | n   |
|---------------------|--------|--------|----------|----------|----------|--------|--------|-----|
| Civil engineering   | 0.1781 | 0.1667 | 0.0143   | 0.6219   | -0.2310  | 0.0052 | 0.5750 | 658 |
| Economics           | 0.1260 | 0.1140 | 0.0105   | 2.0595   | 8.3647   | 0.0083 | 0.7381 | 132 |
| Computer Sciences   | 0.1702 | 0.1528 | 0.0139   | 0.8284   | 0.5582   | 0.0085 | 0.6410 | 255 |
| Control engineering | 0.1715 | 0.1667 | 0.0133   | 0.6833   | -0.1512  | 0.0132 | 0.4884 | 97  |

<https://doi.org/10.1371/journal.pone.0290885.t004>

universities to promote students' success [46]. The fact that many students do not achieve their goals during university experience is a waste of talent and human potential [42, 46].

For modeling these data, we fit five  $UREW$  special models studied in the current paper, i.e., the  $URG$ ,  $URBXII$ ,  $URL$ ,  $URR$ , and  $URW$  distributions. Their densities are given by equations (5), (7), (9), (11), and (10), respectively. We fix  $\tau$  at 0.5 in those equations. We also considered six well-known alternative distributions to describe random variables supported in the unit interval for comparison purposes. We fit the Beta,  $KW$ ,  $UBS$ ,  $UG$ ,  $UW$ , and complementary unit Weibull ( $CUW$ ) [14] distributions. They do not represent  $UREW$  special cases and are selected as competitive distributions due to their relevance in the literature. The beta and  $KW$  are classical models for double-bounded outcomes. The  $UG$  is chosen due to its relevance to various problems. It has received a great deal of attention from statisticians for developing methodological advances [47]. The  $UBS$  and  $UW$  are two of the most relevant models regarding recent advances in distribution theory. The  $CUW$  arises as an alternative model due to its usefulness regarding educational modeling. This distribution has proved helpful in analyzing literacy rates [14]. The densities of all these competitor models are presented in Appendix A.

Parameter estimation is performed by the maximum likelihood method for all fitted models, and the Cramér-von Mises corrected statistic [48] ( $W^*$ ) is considered as the goodness-of-fit measure. Those estimates are computed using the `goodness.fit` function from the `AdequacyModel` package [49]. The `goodness.fit` function allows computing the MLEs of probability distributions and their goodness-of-fit statistical measures. It uses the `optim` function in the implementation and includes several optimization techniques. For the paper results, we use the BFGS algorithm and compute the observed information matrix numerically. Thus, the standard errors and confidence intervals were obtained from the asymptotic normality property of the MLEs. We set the initial values at 1 for the shape (or precision) parameter, the sample mean for the distributions indexed in the mean, and the sample quantile for those with quantile parametrization.

The estimation results for all data sets are reported in Table 5. We observe that the distributions on the  $UREW$  family have the lowest  $W^*$  for the course types considered. The proposed models occupy the first three positions in the ranking for civil engineering and computer sciences. Analyzing the control engineering course, we note that the  $URW$  outperforms the others and is followed by the  $URL$  distribution, which also belongs to the  $UREW$  family. For the economics course, the  $URL$  distribution has superior goodness-of-fit. Fig 4 displays the box-plots and the histograms with fitted density functions for the three best models according to  $W^*$ . Those plots corroborate that the  $UREW$  fits are adequate to the dropout rates of all course types considered and provides real improvement over existing distributions. Therefore, the proposed family is shown competitive with classical unit models such as the beta and  $KW$  distributions.

The  $UREW$  special cases also exhibit superior performance when compared to recent alternatives, including the  $UBS$ ,  $UW$ , and  $CUW$  distributions. It is worth noting that the  $CUW$  distribution has been commonly used in educational modeling. In [14], it was verified that this model can properly fit literacy rates. However, it is important to highlight that while higher literacy rates are desirable [14], lower values are considered more favorable in the case of dropout rates [17]. In this case, it is expected that left-skewed distributions to fit better the former and right-skewed distributions to be more suitable for the latter. This feature may explain why the  $CUW$  is not among the best models for the analyzed datasets while evincing the capacity of the  $UREW$  family to model the first-year dropout rate effectively.

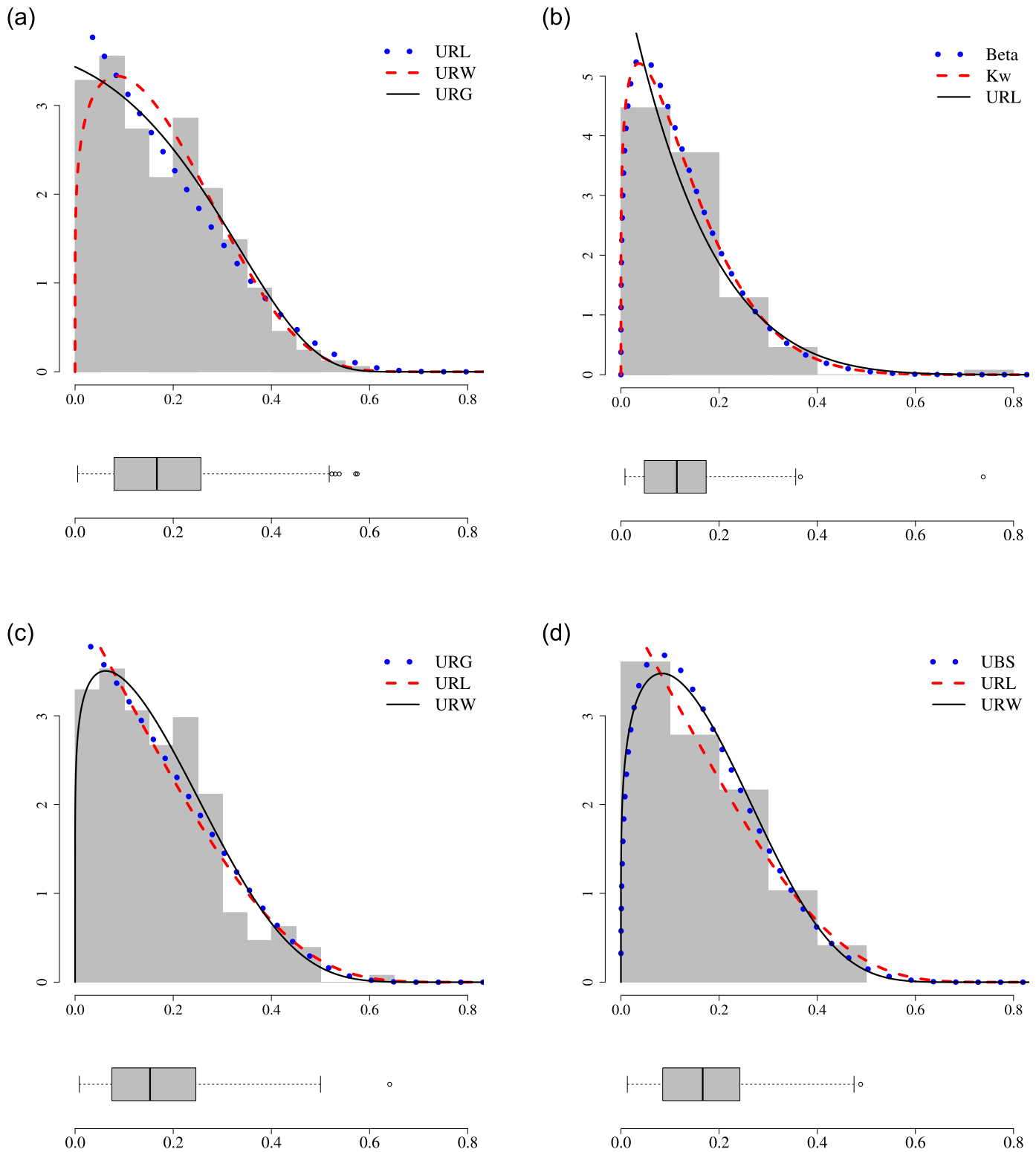
Our results may represent useful tools for universities to evaluate and improve their programs. It is a relevant application as it allows us to deal with the academic, social, and

Table 5. MLEs estimates, the corresponding standard errors (given in parentheses) and goodness-of-fit measure for all fitted models and course types considered.

| Distributions          | Civil engineering |          |        |         | Economics           |          |        |         |
|------------------------|-------------------|----------|--------|---------|---------------------|----------|--------|---------|
|                        | Estimates         |          | W*     | Ranking | Estimates           |          | W*     | Ranking |
| Beta( $\varphi, \mu$ ) | 8.7209            | 0.1771   | 0.4402 | 9       | 10.8025             | 0.1278   | 0.0691 | 2       |
|                        | (0.4763)          | (0.0048) |        |         | (1.3524)            | (0.0084) |        |         |
| KW( $\phi, \mu$ )      | 1.3856            | 0.1553   | 0.3650 | 8       | 1.2090              | 0.1052   | 0.0706 | 3       |
|                        | (0.0494)          | (0.0049) |        |         | (0.0895)            | (0.0085) |        |         |
| UG( $\phi, \mu$ )      | 5.569             | 0.1771   | 0.5817 | 10      | 6.8278              | 0.1265   | 0.0787 | 7       |
|                        | (0.2983)          | (0.0049) |        |         | (0.8207)            | (0.0087) |        |         |
| UBS( $\beta, \mu$ )    | 0.4411            | 1.8501   | 0.2562 | 4       | 0.4238              | 2.2141   | 0.0877 | 8       |
|                        | (0.0122)          | (0.0310) |        |         | (0.0261)            | (0.0798) |        |         |
| UNW( $\beta, \mu$ )    | 2.4602            | 0.1382   | 1.3007 | 11      | 2.8739              | 0.0920   | 0.1929 | 10      |
|                        | (0.0719)          | (0.0051) |        |         | (0.1902)            | (0.0078) |        |         |
| UNW( $\beta, \mu$ )    | 1.3224            | 0.1572   | 0.3128 | 5       | 1.1478              | 0.1043   | 0.0712 | 5       |
|                        | (0.0409)          | (0.0050) |        |         | (0.0723)            | (0.0088) |        |         |
| URBXII( $\beta, \mu$ ) | 1.2876            | 0.1576   | 0.3167 | 6       | 1.1415              | 0.1049   | 0.072  | 6       |
|                        | (0.0364)          | (0.0050) |        |         | (0.0670)            | (0.0088) |        |         |
| URG( $\beta, \mu$ )    | 0.7173            | 0.1586   | 0.1719 | 1       | 0.7471              | 0.0947   | 0.0710 | 4       |
|                        | (0.1731)          | (0.0059) |        |         | (0.3012)            | (0.0084) |        |         |
| URL( $\beta, \mu$ )    | 50.2648           | 0.1456   | 0.2462 | 3       | 1.111               | 0.0955   | 0.0687 | 1       |
|                        | (54.5751)         | (0.0049) |        |         | (0.6462)            | (0.0087) |        |         |
| URR( $\mu$ )           | 0.2133            |          | 0.3412 | 7       | 0.2074              |          | 0.3673 | 11      |
|                        | (0.0033)          |          |        |         | (0.0072)            |          |        |         |
| URW( $\beta, \mu$ )    | 1.1797            | 0.1609   | 0.2117 | 2       | 0.9861              | 0.1047   | 0.0936 | 9       |
|                        | (0.0358)          | (0.0053) |        |         | (0.0585)            | (0.0096) |        |         |
| Distributions          | Computer sciences |          |        |         | Control engineering |          |        |         |
|                        | Estimates         |          | W*     | Ranking | Estimates           |          | W*     | Ranking |
| Beta( $\varphi, \mu$ ) | 8.689             | 0.1695   | 0.1371 | 8       | 9.1876              | 0.1706   | 0.0494 | 8       |
|                        | (0.7654)          | (0.0075) |        |         | (1.3106)            | (0.0119) |        |         |
| KW( $\phi, \mu$ )      | 1.3346            | 0.1466   | 0.1038 | 7       | 1.3969              | 0.1496   | 0.0371 | 7       |
|                        | (0.0761)          | (0.0076) |        |         | (0.1285)            | (0.0121) |        |         |
| UG( $\phi, \mu$ )      | 5.5231            | 0.1695   | 0.215  | 10      | 5.8359              | 0.1707   | 0.0756 | 10      |
|                        | (0.4751)          | (0.0078) |        |         | (0.8152)            | (0.0124) |        |         |
| UBS( $\beta, \alpha$ ) | 0.4455            | 1.9038   | 0.071  | 4       | 0.4299              | 1.8902   | 0.0294 | 3       |
|                        | (0.0197)          | (0.0518) |        |         | (0.0309)            | (0.0806) |        |         |
| UNW( $\beta, \mu$ )    | 2.4404            | 0.1304   | 0.5647 | 11      | 2.5101              | 0.133    | 0.2082 | 11      |
|                        | (0.1141)          | (0.0080) |        |         | (0.1906)            | (0.0128) |        |         |
| UNW( $\beta, \mu$ )    | 1.2758            | 0.148    | 0.0857 | 6       | 1.3319              | 0.1514   | 0.0315 | 5       |
|                        | (0.0626)          | (0.0079) |        |         | (0.1068)            | (0.0125) |        |         |
| URBXII( $\beta, \mu$ ) | 1.2489            | 0.1485   | 0.0833 | 5       | 1.2956              | 0.1518   | 0.0309 | 4       |
|                        | (0.0561)          | (0.0079) |        |         | (0.0949)            | (0.0126) |        |         |
| URG( $\beta, \mu$ )    | 0.2618            | 0.1446   | 0.0564 | 2       | 0.8688              | 0.1532   | 0.0318 | 6       |
|                        | (0.2490)          | (0.0088) |        |         | (0.4991)            | (0.0148) |        |         |
| URL( $\beta, \mu$ )    | 28.8352           | 0.1396   | 0.0673 | 3       | 33.445              | 0.1394   | 0.0282 | 2       |
|                        | (51.2034)         | (0.0076) |        |         | (76.399)            | (0.0123) |        |         |
| URR( $\mu$ )           | 0.2111            |          | 0.1705 | 9       | 0.2038              |          | 0.0676 | 9       |
|                        | (0.0052)          |          |        |         | (0.0082)            |          |        |         |
| URW( $\beta, \mu$ )    | 1.1321            | 0.1512   | 0.0561 | 1       | 1.1936              | 0.1547   | 0.0258 | 1       |
|                        | (0.0541)          | (0.0083) |        |         | (0.0942)            | (0.0131) |        |         |

<https://doi.org/10.1371/journal.pone.0290885.t005>





**Fig 4. Histogram and estimated densities for the applications.**

<https://doi.org/10.1371/journal.pone.0290885.g004>

economic implications of university dropout [17]. Nevertheless, other potential applications can be explored in the context of educational modeling. The new family can be competitive to model literacy rates [14], educational attainment percentages [31], proportions of adolescents who want top grades at school [32], and proportions of the novice teachers with a mentor at the school [33]. These variables have been explored through other commonly used distributions in educational modeling. We can also cite the graduation and persistence rates as further applications, which are related to student progression and academic success patterns [34].

## 7 Final remarks

This paper defines the unit ratio-extended Weibull ( $UR\mathcal{E}\mathcal{W}$ ) family of distributions. It is obtained on a ratio transformation in the extended Weibull family and can be used to model continuous random variables in the unit interval. The new family has a closed-form for quantile measures; thus, we provide a quantile parametrization for the family. Several special cases are derived, and parameter estimation is explored using the maximum likelihood theory. We show that some one-parameter  $UR\mathcal{E}\mathcal{W}$  special cases may present closed-form for the maximum likelihood estimator (MLE). We perform Monte Carlo experiments to assess the performance of those estimators. For example, the unit ratio-Rayleigh MLE is approximately unbiased for small sample sizes. We also note an appropriate performance for the unit ratio-Gompertz MLEs. The utility of the proposed family is illustrated with applications to the first-year dropout rate of undergraduate courses in Brazilian universities. We select four course types and note that, for those data, the  $UR\mathcal{E}\mathcal{W}$  special models fit properly and outperform other classical and recent unit distributions. Thus, the new family can be competitive alternative when those models are unsuitable. We emphasize that a long list of possibilities can be addressed in future works. For example, our approach can be investigated in the presence of zeros and ones, and quantile regression models are also a natural path. The  $UR\mathcal{E}\mathcal{W}$  can also be generalized to accommodate time-dependent double-bounded indicators by using the autoregressive moving average models. This kind of structure is in the state-of-art literature on the analysis of double-bounded time series. The  $UR\mathcal{E}\mathcal{W}$  can also attract applications to other double-bounded variables, being a competitive option to other unit distributions commonly used in educational modeling. For instance, literacy rates, educational attainment percentages, graduation, and persistence rates are educational measurements that represent potential applications for the proposed family.

## Appendix

### A—Alternative distributions fitted in the applications

In this appendix, we present the unit distributions fitted in Section 6 as alternative models to the  $UR\mathcal{E}\mathcal{W}$  family. These model and their corresponding densities are listed bellow (for  $0 < y < 1$ ):

- The beta density is given by

$$f_{\text{Beta}}(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1},$$

where  $\mu \in (0, 1)$  is the mean of  $Y$  and  $\phi > 0$  is a precision parameter. The above parametrization is pioneered by [50].

- The  $\mathcal{KW}$  density is given by

$$f_{\mathcal{KW}}(y) = \frac{\phi \log(1 - q)}{\log(1 - e^{-\phi/\log(\mu)})} y^{-\phi/\log(\mu)-1} (1 - y^{-\phi/\log(\mu)})^{\log(1-q)/\log(1-e^{-\phi})-1},$$

where  $\mu \in (0, 1)$  is the  $q$ th quantile parameter, and  $\phi > 0$  is a precision parameter. The above parametrization is pioneered by [51]. In Section 6 we fix  $q$  at 0.5 thus the parameter  $\mu$  refers to the median of  $Y$ .

- The  $\mathcal{UG}$  density is given by

$$f_{\mathcal{UG}}(y) = \left[ \frac{\mu^{1/\phi}}{1 - \mu^{1/\phi}} \right]^\phi \frac{1}{\Gamma(\phi)} y^{\mu^{1/\phi}/[(1-\mu^{1/\phi})-1]} [-\log(y)]^{\phi-1},$$

where  $\mu \in (0, 1)$  is the mean of  $Y$  and  $\phi > 0$  is a precision parameter. The above parametrization is pioneered by [52].

- The  $\mathcal{UBS}$  density is given by

$$f_{\mathcal{UBS}}(y) = \frac{1}{2y\alpha\beta\sqrt{2\pi}} \left[ \left( -\frac{\beta}{\log(y)} \right)^{\frac{1}{2}} + \left( -\frac{\beta}{\log(y)} \right)^{\frac{3}{2}} \right] \times \exp \left\{ \frac{1}{2\alpha^2} \left( \frac{\log(y)}{\beta} + \frac{\beta}{\log(y)} + 2 \right) \right\},$$

where  $\alpha > 0$  and  $\beta > 0$  are shape parameters. The  $\mathcal{UBS}$  is pioneered by [11].

- The  $\mathcal{UW}$  density is given by

$$f_{\mathcal{UW}}(y) = \frac{\beta}{y} \left( \frac{\log \tau}{\log \mu} \right) \left( \frac{\log y}{\log \mu} \right)^{\beta-1} \tau^{(\log y/\log \mu)^\beta},$$

where  $\mu \in (0, 1)$  is the  $\tau$ th quantile parameter and  $\beta > 0$  are shape parameters. The above parametrization is pioneered by [41]. In Section 6 we fix  $\tau$  at 0.5 thus the parameter  $\mu$  refers to the median of  $Y$ .

- The  $\mathcal{CUW}$  density is given by

$$f_{\mathcal{CUW}}(y) = \frac{\beta \log 2}{1 - y} [-\log(1 - \mu)]^{-\beta} [-\log(1 - y)]^{\beta-1} 2^{-[\log(1-y)/\log(1-\mu)]^\beta},$$

where  $\mu \in (0, 1)$  is the median of  $Y$  and  $\beta > 0$  is a shape parameter. The above distribution is pioneered by [14].

## Supporting information

**S1 Data.**  
(ZIP)

## Author Contributions

**Conceptualization:** Fernando A. Peña-Ramírez, Renata R. Guerra.

**Data curation:** Renata R. Guerra.

**Formal analysis:** Renata R. Guerra.

**Investigation:** Fernando A. Peña-Ramírez, Charles Peixoto Mafalda.

**Methodology:** Renata R. Guerra.

**Project administration:** Fernando A. Peña-Ramírez.

**Resources:** Fernando A. Peña-Ramírez.

**Software:** Renata R. Guerra, Charles Peixoto Mafalda.

**Supervision:** Fernando A. Peña-Ramírez.

**Validation:** Fernando A. Peña-Ramírez.

**Visualization:** Renata R. Guerra.

**Writing – original draft:** Fernando A. Peña-Ramírez, Renata R. Guerra, Charles Peixoto Mafalda.

**Writing – review & editing:** Fernando A. Peña-Ramírez.

## References

1. Tahir MH, Nadarajah S. Parameter induction in continuous univariate distributions: Well-established G families. *Annals of the Brazilian Academy of Sciences*. 2015; 87:539–568. <https://doi.org/10.1590/0001-3765201520140299> PMID: 26131628
2. Peña-Ramírez FA, Guerra RR, Cordeiro GM, Marinho PR. The exponentiated power generalized Weibull: Properties and applications. *Anais da Academia Brasileira de Ciências*. 2018; 90:2553–2577. <https://doi.org/10.1590/0001-3765201820170423> PMID: 30304207
3. Zichuan M, Hussain S, Iftikhar A, Ilyas M, Ahmad Z, Khan DM, et al. A new extended-family of distributions: properties and applications. *Computational and Mathematical Methods in Medicine*. 2020; 2020. <https://doi.org/10.1155/2020/4650520> PMID: 32549906
4. Arif M, Khan DM, Khosa SK, Aamir M, Aslam A, Ahmad Z, et al. Modelling insurance losses with a new family of heavy-tailed distributions. *Computers, Materials & Continua*. 2021; 66:537–550. <https://doi.org/10.32604/cmc.2020.012420>
5. Peña-Ramírez FA, Guerra RR, Cordeiro GM. The Nadarajah-Haghighi Lindley distribution. *Anais da Academia Brasileira de Ciências*. 2019; 91:e20170856. <https://doi.org/10.1590/0001-3765201920170856> PMID: 30994747
6. Peña-Ramírez FA, Guerra RR, Canterle DR, Cordeiro GM. The logistic Nadarajah–Haghighi distribution and its associated regression model for reliability applications. *Reliability Engineering & System Safety*. 2020; 204:107196. <https://doi.org/10.1016/j.res.2020.107196>
7. Kumaraswamy P. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*. 1980; 46:79–88. [https://doi.org/10.1016/0022-1694\(80\)90036-0](https://doi.org/10.1016/0022-1694(80)90036-0)
8. Grassia A. On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma distribution and derived compound distributions. *Australian Journal of Statistics*. 1977; 19:108–114. <https://doi.org/10.1111/j.1467-842X.1977.tb01277.x>
9. Barndorff-Nielsen OE, Jorgensen B. Some parametric models on the simplex. *Journal of Multivariate Analysis*. 1991; 39:106–116. [https://doi.org/10.1016/0047-259X\(91\)90008-P](https://doi.org/10.1016/0047-259X(91)90008-P)
10. Smithson M, Shou Y. CDF-quantile distributions for modelling random variables on the unit interval. *British Journal of Mathematical and Statistical Psychology*. 2017; 70:412–438. <https://doi.org/10.1111/bmsp.12091> PMID: 28306155
11. Mazucheli J, Menezes AFB, Dey S. The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*. 2018; 9:47–57.
12. Mazucheli J, Leiva V, Alves B, Menezes AF. A new quantile Regression for modeling bounded data under a unit Birnbaum–Saunders distribution with applications in medicine and politics. *Symmetry*. 2021; 13:682. <https://doi.org/10.3390/sym13040682>
13. Mazucheli J, Menezes A, Ghitany M. The unit-Weibull distribution and associated inference. *Journal of Applied Probability and Statistics*. 2018; 13:1–22.
14. Guerra RR, Peña-Ramírez FA, Bourguignon M. The unit extended Weibull families of distributions and its applications. *Journal of Applied Statistics*. 2021; 48:3174–3192. <https://doi.org/10.1080/02664763.2020.1796936> PMID: 35707261

15. Mazucheli J, Menezes AF, Dey S. Unit-Gompertz distribution with applications. *Statistica*. 2019; 79:25–43.
16. Korkmaz MÇ, Chesneau C. On the unit Burr-XII distribution with the quantile regression modeling and applications. *Computational and Applied Mathematics*. 2021; 40:1–26. <https://doi.org/10.1007/s40314-021-01418-5>
17. Ribeiro TF, Peña-Ramírez FA, Guerra RR, Cordeiro GM. Another unit Burr XII quantile regression model based on the different reparameterization applied to dropout in Brazilian undergraduate courses. *Plos one*. 2022; 17:e0276695. <https://doi.org/10.1371/journal.pone.0276695> PMID: 36327245
18. Ribeiro TF, Cordeiro GM, Pena-Ramirez FA, Guerra RR. A new quantile regression for the COVID-19 mortality rates in the United States. *Computational and Applied Mathematics*. 2021; 40:1–16. <https://doi.org/10.1007/s40314-021-01553-z>
19. Korkmaz MÇ. The unit generalized half normal distribution: A new bounded distribution with inference and application. *UPB Scientific Bulletin, Series A: Applied Mathematics and Physics*. 2020; 82:133–140.
20. Nasiru S, Abubakari AG, Angbing ID. Bounded odd inverse pareto exponential distribution: Properties, estimation, and regression. *International Journal of Mathematics and Mathematical Sciences*. 2021; 2021:1–18. <https://doi.org/10.1155/2021/9955657>
21. Sagrillo M, Guerra RR, Bayer FM. Modified Kumaraswamy distributions for double bounded hydro-environmental data. *Journal of Hydrology*. 2021; 603:127021. <https://doi.org/10.1016/j.jhydrol.2021.127021>
22. Martínez-Flórez G, Tovar-Falón R. New regression models based on the unit-sinh-normal distribution: Properties, inference, and applications. *Mathematics*. 2021; 9:1231. <https://doi.org/10.3390/math9111231>
23. Altun E, El-Morshedy M, Eliwa M. A new regression model for bounded response variable: An alternative to the beta and unit-Lindley regression models. *Plos one*. 2021; 16:e0245627. <https://doi.org/10.1371/journal.pone.0245627> PMID: 33481884
24. Altun E. The log-weighted exponential regression model: alternative to the beta regression model. *Communications in Statistics—Theory and Methods*. 2021; 50:2306–2321. <https://doi.org/10.1080/03610926.2019.1664586>
25. Espinheira PL, Silva LCM, Cribari-Neto F. Bias and variance residuals for machine learning nonlinear simplex regressions. *Expert Systems with Applications*. 2021; 185:115656. <https://doi.org/10.1016/j.eswa.2021.115656>
26. Yero EJH, Sacco NC, do Carmo Nicoletti M. Effect of the municipal Human Development index on the results of the 2018 Brazilian presidential Elections. *Expert Systems with Applications*. 2020; 168:114305. <https://doi.org/10.1016/j.eswa.2020.114305>
27. Bayer FM, Bayer DM, Pumi G. Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*. 2017; 555:385–396. <https://doi.org/10.1016/j.jhydrol.2017.10.006>
28. Melchior C, Zanini RR, Guerra RR, Rockenbach DA. Forecasting Brazilian mortality rates due to occupational accidents using autoregressive moving average approaches. *International Journal of Forecasting*. 2021; 37:825–837. <https://doi.org/10.1016/j.ijforecast.2020.09.010>
29. Calabrese R, Zanin L. Modelling spatial dependence for loss given default in peer-to-peer lending. *Expert Systems with Applications*. 2022; 192:116295. <https://doi.org/10.1016/j.eswa.2021.116295>
30. Gurvich M, DiBenedetto A, Ranade S. A new statistical distribution for characterizing the random strength of brittle materials. *Journal of Materials Science*. 1997; 32:2559–2564. <https://doi.org/10.1023/A:1018603118573>
31. Korkmaz M, Chesneau C, Korkmaz ZS. Transmuted unit Rayleigh quantile regression model: Alternative to beta and Kumaraswamy quantile regression models. *Univ Politeh Buchar Sci Bull Ser Appl Math Phys*. 2021; 83:149–158.
32. Korkmaz MÇ, Chesneau C, Korkmaz ZS. A new alternative quantile regression model for the bounded response with educational measurements applications of OECD countries. *Journal of Applied Statistics*. 2023; 50:131–154. <https://doi.org/10.1080/02664763.2021.1981834> PMID: 36530782
33. Korkmaz MÇ, Korkmaz ZS. The unit log–log distribution: a new unit distribution with alternative quantile regression modeling and educational measurements applications. *Journal of Applied Statistics*. 2021; p. 1–20.
34. Sneyers E, De Witte K. The interaction between dropout, graduation rates and quality ratings in universities. *Journal of the Operational Research Society*. 2017; 68:416–430. <https://doi.org/10.1057/jors.2016.15>
35. Cave M. The use of performance indicators in higher education: A critical analysis of developing practice. *Higher Education Policy Series, 2*. ERIC; 1991.

36. Marinho PRD, Cordeiro GM, Ramírez FP, Alizadeh M, Bourguignon M. The exponentiated logarithmic generated family of distributions and the evaluation of the confidence intervals by percentile bootstrap. *Brazilian Journal of Probability and Statistics*. 2018; 32:281–308. <https://doi.org/10.1214/16-BJPS343>
37. Abbas K, Hussain Z, Rashid N, Ali A, Taj M, Khan SA, et al. Bayesian estimation of Gumbel type-II distribution under type-II censoring with medical applications. *Computational and Mathematical Methods in Medicine*. 2020; 2020:1–11. <https://doi.org/10.1155/2020/1876073>
38. Peña-Ramírez FA, Guerra RR, Cordeiro GM. A new Nadarajah-Haghighi generalization with five different shapes for the hazard function. *Revista Colombiana de Estadística*. 2023; 46:1–29.
39. Mitnik PA, Baek S. The Kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation. *Statistical Papers*. 2013; 54:177–192. <https://doi.org/10.1007/s00362-011-0417-y>
40. Lemonte AJ, Bazán JL. New class of Johnson distributions and its associated regression model for rates and proportions. *Biometrical Journal*. 2016; 58:727–746. <https://doi.org/10.1002/bimj.201500030> PMID: 26659998
41. Mazucheli J, Menezes A, Fernandes L, De Oliveira R, Ghitany M. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. *Journal of Applied Statistics*. 2020; 47:954–974. <https://doi.org/10.1080/02664763.2019.1657813> PMID: 35706917
42. Ferrão ME, Almeida LS. Multilevel modeling of persistence in higher education. *Ensaio: Avaliação e Políticas Públicas em Educação*. 2018; 26:664–683.
43. Thammasiri D, Delen D, Meesad P, Kasap N. A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*. 2014; 41:321–330. <https://doi.org/10.1016/j.eswa.2013.07.046>
44. Sneyers E, De Witte K. Interventions in higher education and their effect on student success: a meta-analysis. *Educational Review*. 2018; 70:208–228. <https://doi.org/10.1080/00131911.2017.1300874>
45. INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: Censo da Educação Superior; 2018. Brasília: Ministério da Educação. Available from: <http://portal.inep.gov.br/basicas-levantamentos-acessar>.
46. Ferrão M, Almeida L. Differential effect of university entrance score on first-year students' academic performance in Portugal. *Assessment & Evaluation in Higher Education*. 2019; 44:610–622. <https://doi.org/10.1080/02602938.2018.1525602>
47. Guedes AC, Cribari-Neto F, Espinheira PL. Modified likelihood ratio tests for unit gamma regressions. *Journal of Applied Statistics*. 2020; 47:1562–1586. <https://doi.org/10.1080/02664763.2019.1683152> PMID: 35707584
48. Chen G, Balakrishnan N. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*. 1995; 27:154–161. <https://doi.org/10.1080/00224065.1995.11979578>
49. Marinho PRD, Silva RB, Bourguignon M, Cordeiro GM, Nadarajah S. AdequacyModel: An R package for probability distributions and general purpose optimization. *PloS one*. 2019; 14:e0221487. <https://doi.org/10.1371/journal.pone.0221487> PMID: 31450236
50. Ferrari SLP, Cribari-Neto F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*. 2004; 7:799–815.
51. Bayes CL, Bazán JL, De Castro M. A quantile parametric mixed regression model for bounded response variables. *Statistics and its interface*. 2017; 10:483–493.
52. Mousa AM, El-Sheikh AA, Abdel-Fattah MA. A gamma regression for bounded continuous variables. *Advances and Applications in Statistics*. 2016; 49:305–326.