REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 491–507
https://doi.org/10.57805/revstat.v21i4.431

# Bias Reduction of Maximum Likelihood Estimates for an Asymmetric Class of Power Models with Applications

Authors: Yolanda M. Gómez
– Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile
ygomez@ubiobio.cl

Bruno Santos
– School of Mathematics Statistics and Actuarial Science, University of Kent, Canterbury, England
b.santos@kent.ac.uk

Diego I. Gallardo
– Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile
dgallardo@ubiobio.cl

Osvaldo Venegas ✉
– Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco, Chile
ovenegas@uct.cl

Héctor W. Gómez
– Departamento de Estadística y Ciencias de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile
hector.gomez@uantof.cl

Received: November 2020 Revised: February 2022 Accepted: February 2022

Abstract:

- In this paper we study some methods to reduce the bias for maximum likelihood estimation in the general class of alpha power models, specifically for the shape parameter. We find the modified maximum likelihood estimator using Firth's method and we show that this estimator is the uniformly minimum variance unbiased estimator (UMVUE) in this class. We consider three special cases of this class, namely the exponentiated exponential (EE), the power half-normal and the power piecewise exponential models. We compare the bias in simulation studies and find that the modified method is definitely superior, especially for small sample sizes, in both the bias and the root mean squared error. We illustrate our modified estimator in four real data set examples, in each of which the modified estimates better explain the variability.

Keywords:

- *UMVUE; Firth's method; exponentiated exponential model; power half-normal model.*

AMS Subject Classification:

- 62E20, 62F10.

✉ Corresponding author.

## 1. INTRODUCTION

The power ($P$) model for a probability density function (pdf), say $f$, was discussed by Lehmann [18] for the case $\alpha \in \mathbb{N}$ and later by Durrans [5] for $\alpha \in \mathbb{R}^+$, the shape parameter. If a random variable $Z$ is distributed according to this class of distributions, we use the notation $Z \sim P_f(\alpha)$.

Its probability density function (pdf) is given by

$$g(z;\alpha) = \alpha f(z) F(z)^{\alpha-1}, \qquad z \in \mathcal{Z}, \tag{1.1}$$

where $\mathcal{Z}$ is the sample space defined for $f$ and $F$ is the cumulative distribution function (cdf) related to $f$. Durrans [5] considers $f = \phi(\cdot)$, the density function of the normal distribution with location and scale parameters, which we will refer to as power normal (PN) distribution. The case when $f$ is symmetrical is discussed by Gupta and Gupta [10], including some fundamental properties of this family; Pewsey *et al.* [38] show that the information matrix is not singular when the symmetrical case is recovered ($\alpha = 1$).

Based on this representation, there are some extensions in the literature and we refer to Table 1 for a few references.

**Table 1**: Some extensions for the Power family.

| **Authors** | **Distribution** | $F$ |
|---|---|---|
| Gupta and Kundu [11, 12] | EE | Exponential |
| Mudholkar *et al.* [32, 33] | Exponentiated Weibull (WE) | Weibull |
| Gupta *et al.* [9] | Exponentiated-Pareto (EP) | Pareto |
| Nadarajah [35] | Exponentiated Gumbel (EG) | Gumbel |
| Kakde and Shirke [14] | Exponentiated lognormal (ELN) | Lognormal |
| Nadarajah and Gupta [36] | Exponentiated gamma (EG) | Gamma |
| Martínez-Flórez *et al.* [25] | Skew-normal alpha power (SNAP) | Skew-normal |
| Martínez-Flórez *et al.* [26] | Power Birnbaum-Saunders (PBS) | Birbaum-Saunders |
| Gómez and Bolfarine [7] | Power half-normal (PHN) | Half-normal |
| Zhao and Kim [42] | Power t (PT) | t-Student |
| Gómez *et al.* [8] | Power piecewise exponential (PPE) | Piecewise exponential |

Another important property of the $P_f(\alpha)$ class is its interpretability for $\alpha \in \mathbb{N}$. In this case, the $P_f(\alpha)$ model can be interpreted as the distribution of the maximum of $X_1, X_2, ..., X_\alpha$, where the $X_i$'s are independent, identically distributed random variables from $X \sim f$.

A similar interpretation is given by Durrans [5] for the extended case $\alpha \in \mathbb{R}^+$ using fractional order statistics.

Other extensions related to this model are presented in Martínez-Flórez *et al.* [23], introducing a multivariate version of the model; Martínez-Flórez *et al.* [24], performing applications in regression models; Martínez-Flórez *et al.* [27], studying the exponential transformation of the model; in [28], studying a version of the doubly censored model with inflation in a regression context. In one of the references in Table 1, Gupta and Kundu [12] reported a simulation study for the EE model in which an overestimation problem for the shape parameter in small sample sizes was observed. Based on this, we propose a bias correction methodology which should be useful not only for this distribution, but for the whole $P$ family. To motivate our discussion, we start our presentation with two special cases from this class of models. Later, in the simulation studies, we add another case; and in the applications we also add another member of this family. This is done to ensure that our bias correction method works with different distributions, not just certain carefully selected distributions.

The first case is the EE model with shape and rate parameters $\alpha$ and $\lambda$ respectively (which we denote as $\mathrm{EE}(\alpha, \lambda)$). The second case is the PHN model with shape and scale parameters $\alpha$ and $\sigma$ (which we denote as $\mathrm{PHN}(\alpha, \sigma)$).

For $Z \sim \mathrm{EE}(\alpha, \lambda)$, the pdf is

$$g(z; \alpha, \lambda) = \alpha\lambda e^{-\lambda z}[1 - e^{-\lambda z}]^{\alpha-1}, \qquad z > 0, \tag{1.2}$$

and for $Z \sim \mathrm{PHN}(\alpha, \sigma)$, the pdf is given by

$$g(z; \alpha, \sigma) = \frac{2\alpha}{\sigma}\phi\Big(\frac{z}{\sigma}\Big)\Big[2\Phi\Big(\frac{z}{\sigma}\Big) - 1\Big]^{\alpha-1}, \qquad z > 0, \tag{1.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and the cdf for the standard normal distribution.

When dealing with these cases, it is important to study the associated bias in parameter estimation, if we use maximum likelihood estimation methods for instance. Although the unbiased property of these estimators is well known in asymptotic condition, we need to be careful when using this estimation method for small sample sizes. For applications of recent mean bias reduction methodology in different contexts, see for instance Kosmidis *et al.* [15], Melo *et al.* [31], Maity *et al.* [22], Magalhães *et al.* [20] and Mazucheli *et al.* [30]. An alternative method is the median bias reduction methodology recently proposed in Pagui *et al.* [37] and applied in different contexts in Kyriakou *et al.* [16] and Ioannis *et al.* [13], for example.

For the $P_f(\alpha)$ class of distributions, a simple simulation study can be set up to identify some weaknesses of the maximum likelihood estimators (MLE's) for different values of each parameter and different sample sizes, considering the cases of the exponentiated exponential and the power-half normal models. In Figures 1 and 2 we report the estimated bias based on 10,000 replicates for the EE and PHN models for the MLEs.

Note that in both models, the observed average bias of the estimator of $\alpha$ is considerably greater (in relative terms) when compared with the average bias for the other parameters. This fact motivates the study of a method for reducing the mean bias for the MLE of $\alpha$ in the general class of model defined in (1.1), which can be applied to any member of the class.

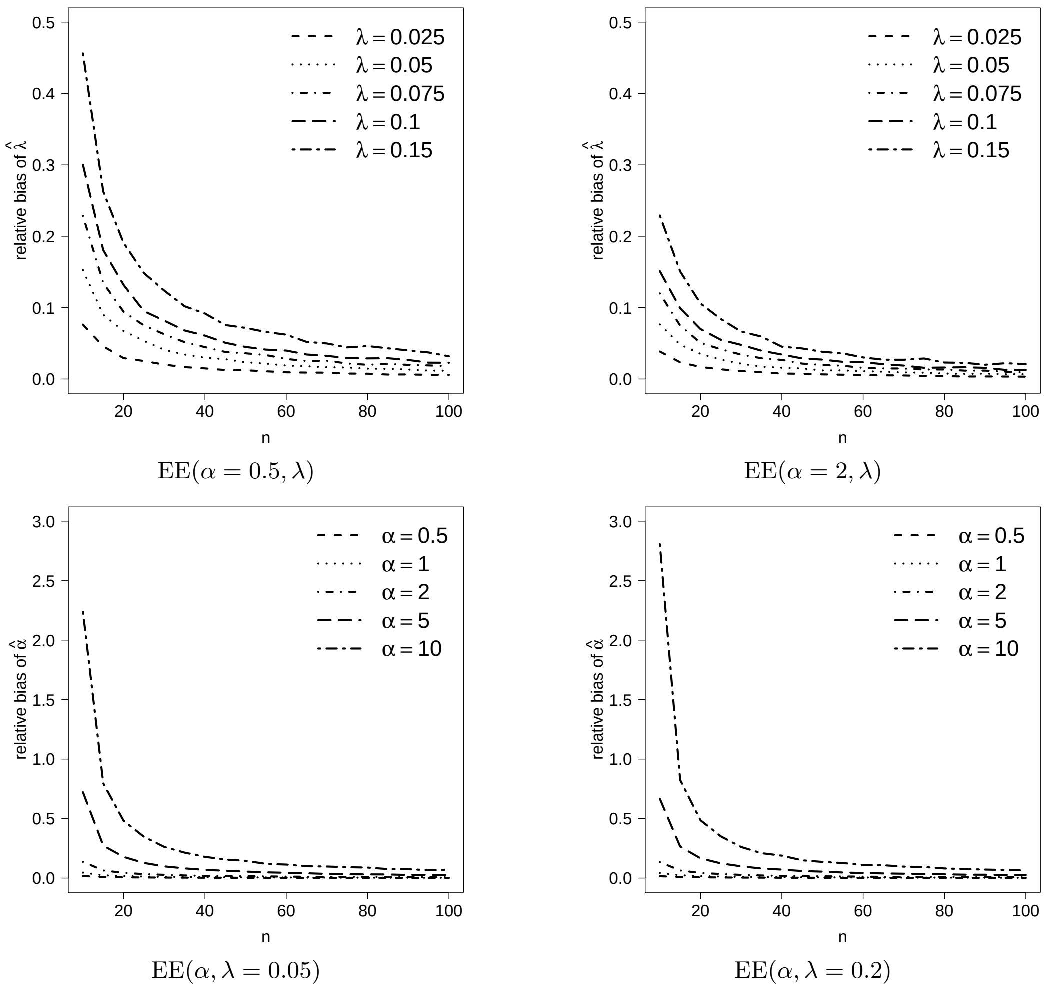



**Figure 1**: Estimated bias for the MLE of $\lambda$ and $\alpha$ in the $\text{EE}(\alpha, \lambda)$ under different scenarios based on 10,000 replicates.

Initially, considering only the shape parameter $\alpha$, it is possible not only to study the bias, but also to define an unbiased estimator, namely the UMVUE. This result is also related to the method proposed by Firth [6] for prevention of bias in maximum likelihood estimation. Moreover, when there are more parameters in this class of distributions, the approach used in Sartori [41] for skew normal models provides a convenient scheme to focus our attention on the bias for $\alpha$, while also maximizing the likelihood for the other parameters; it has been applied more recently in Arrué *et al.* [1] and Magalhães *et al.* [21] in the study of bias for skew-normal, modified skew-normal and Marshall-Olkin models respectively.

This paper is organized as follows. Section 2 defines the bias for the shape parameter and presents the UMVUE for $\alpha$ in the general class of power models. Section 3 discusses the method to prevent bias to the shape parameter, while also obtaining maximum likelihood estimates for the other parameters, where we describe an iterative algorithm to find these estimates. Section 4 shows a simulation study, where we consider the cases not only of the exponentiated exponential and the power-half models, but also the power piecewise exponential, to illustrate the superior performance of the modified estimator. In Section 5, we highlight the improvements provided by the methods proposed here with three applications, which are known in the literature for this type of data.

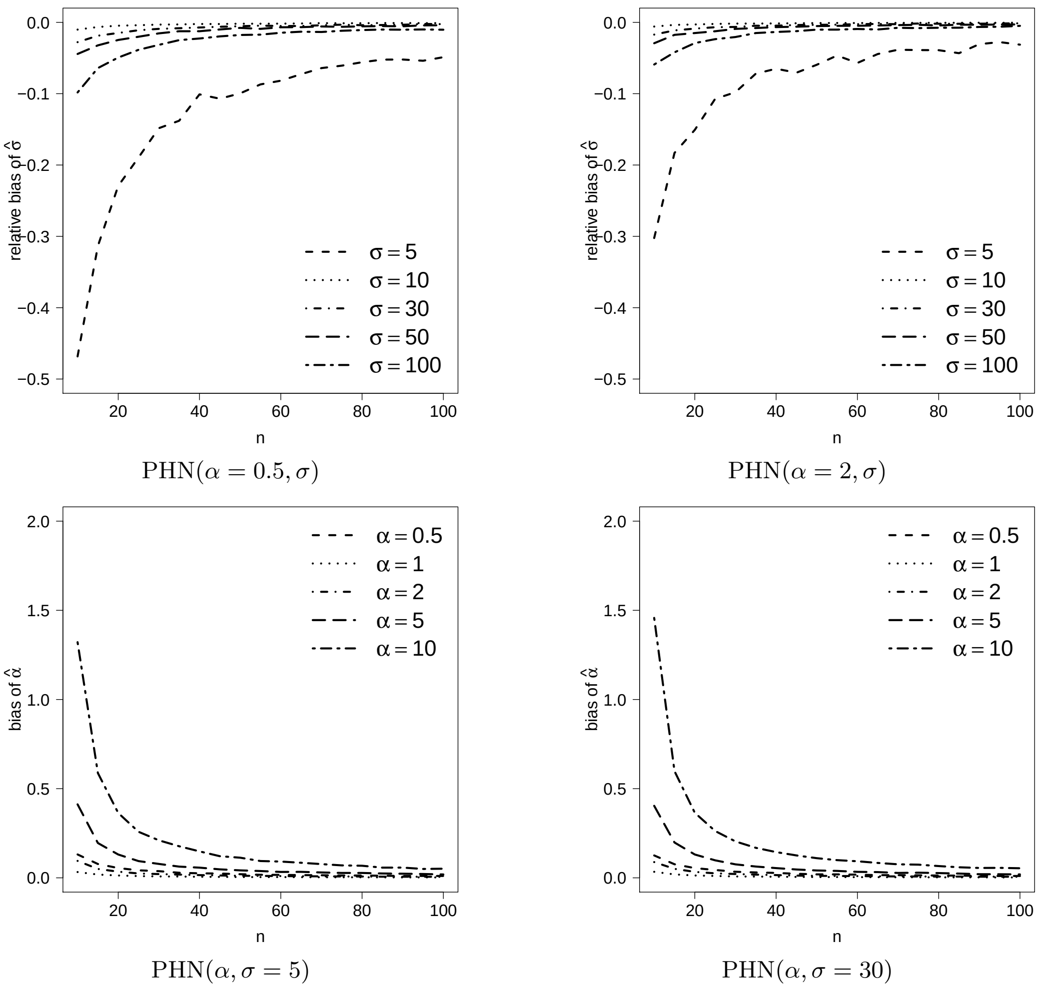



**Figure 2**: Estimated bias for the MLE of $\sigma$ and $\alpha$ in the PHN$(\alpha,\sigma)$ under different scenarios based on 10,000 replicates.

## 2. CASE I: $F(\cdot)$ IS FREE OF PARAMETERS

The likelihood for a random sample $\boldsymbol{z} = (z_1, z_2, ..., z_n)$ from $P_f(\alpha)$ is given by

$$L(\alpha) = \prod_{i=1}^{n} f(z_i) \times \alpha^n \exp\left\{(\alpha-1)\left(\sum_{i=1}^{n} \log F(z_i)\right)\right\}. \tag{2.1}$$

**Theorem 2.1.** *For the $P_f(\alpha)$ model, $T(\boldsymbol{z}) = -\sum_{i=1}^{n} \log F(z_i)$ is a complete statistic.*

**Proof:** Note that the likelihood function in (2.1) can be broken down as

$$L(\alpha) = \underbrace{\prod_{i=1}^{n} f(z_i)}_{h(\boldsymbol{z})} \times \underbrace{\alpha^n \exp\left\{-(\alpha-1)\left(-\sum_{i=1}^{n} \log F(z_i)\right)\right\}}_{g_\alpha\left(-\sum_{i=1}^{n} \log F(z_i)\right)}.$$

According to the Neyman factorization theorem, $T = T(\boldsymbol{z}) = -\sum_{i=1}^{n} \log F(z_i)$ is a sufficient statistic for $\alpha$. It is possible to verify that $-\log F(Z_i) \sim E(\alpha)$, where $E(\alpha)$ denotes the exponential distribution with rate $\alpha$. As the $Z_i$'s are independent, then $T \sim G(n, \alpha)$, with $G(a, b)$ the gamma distribution with shape and rate parameters $a$ and $b$ respectively. Let $g$ be a measurable function in the interval $(0, \infty)$. Therefore,

$$\mathbb{E}(g(T)) = 0 \Leftrightarrow \int_0^{\infty} g(t)t^{n-1}e^{-\alpha t}dt = 0.$$

As $\alpha > 0$ and $n \geq 1$, then $t^{n-1}e^{-\alpha t} > 0$, $\forall t > 0$. Thus, $\mathbb{E}(g(T)) = 0 \Leftrightarrow g(T) = 0$, implying that $T$ is a complete statistic. □

On the other hand, the log-likelihood function is given by

$$\ell(\alpha) = n \log \alpha + \sum_{i=1}^{n} [\log f(z_i) + (\alpha - 1) \log F(z_i)].$$

It is direct that the MLE of $\alpha$ is given by

$$\widehat{\alpha} = \frac{n}{-\sum_{i=1}^{n} \log F(z_i)}.$$

**Theorem 2.2.** *$\widehat{\alpha}$ is a biased estimator for $\alpha$.*

**Proof:** As $-\sum_{i=1}^{n} \log F(Z_i) \sim G(n, \alpha)$, we have that

$$\mathbb{E}(\widehat{\alpha}) = \frac{n\alpha}{n-1}, \qquad n > 1.$$ □

**Remark 2.1.** Note that bias$(\widehat{\alpha}) = \alpha/(n-1)$, so that the bias can be "too large" when $\alpha$ is increased and the sample size is small. Clearly, for $n \to \infty$, $\widehat{\alpha}$ is unbiased.

**Theorem 2.3.** *$\widehat{\alpha}_M = (n-1)/(-\sum_{i=1}^{n} \log F(z_i))$ is the UMVUE for $\alpha$.*

**Proof:** It is clear that $\widehat{\alpha}_M$ is an unbiased estimator for $\alpha$. As $\widehat{\alpha}_M$ depends on a complete statistic, by the Lehmann-Scheffé theorem $\widehat{\alpha}_M$ is the UMVUE for $\alpha$. □

### 2.1. Connection with the Firth method

A popular method to reduce the bias of an estimator is the Firth method [6]. For the univariate case, the method consists in modifying the score function, say $S(\alpha)$, by

$$S_M(\alpha) = S(\alpha) + M(\alpha),$$

where $M(\alpha) = \frac{1}{2} I(\alpha)^{-1}(\nu_{\alpha,\alpha,\alpha} + \nu_{\alpha,\alpha\alpha})$, $I(\alpha)$ the information matrix for the model, $\nu_{\alpha,\alpha,\alpha} = \mathbb{E}\left[\left(\frac{\partial \ell(\alpha)}{\partial \alpha}\right)^3\right]$ and $\nu_{\alpha,\alpha\alpha} = \mathbb{E}\left[\frac{\partial \ell(\alpha)}{\partial \alpha} \frac{\partial^2 \ell(\alpha)}{\partial \alpha^2}\right]$. The solution of the modified score equation $S_M(\alpha) = 0$ produces the modified MLE, say $\widehat{\alpha}_M$. Firth [6] shows that the bias of $\widehat{\alpha}_M$ is reduced from $O(n^{-1})$ to $O(n^{-2})$ when compared with the ordinary MLE. Moreover, the asymptotic distribution of $\widehat{\alpha}_M$ coincides with that of $\widehat{\alpha}$, i.e.

$$\sqrt{n}(\widehat{\alpha}_M - \alpha) \to N\big(0, I(\alpha)^{-1}\big), \quad \text{as } n \to \infty.$$

Note that for the $P_f(\alpha)$ model

$$\frac{\partial \ell(\alpha)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log F(z_i) \qquad \text{and} \qquad \frac{\partial^2 \ell(\alpha)}{\partial \alpha^2} = -\frac{n^2}{\alpha^2}.$$

As $-\sum_{i=1}^{n} \log F(Z_i) \sim G(n, \alpha)$, it can be verified that

$$I(\alpha) = \frac{n^2}{\alpha^2}, \quad \nu_{\alpha,\alpha,\alpha} = -\frac{2n}{\alpha^3} \quad \text{and} \quad \nu_{\alpha,\alpha\alpha} = 0.$$

Then, for the $P_f(\alpha)$ model we have that $M(\alpha) = -\alpha^{-1}$. Therefore, $S_M(\alpha) = \frac{n-1}{\alpha} + \sum_{i=1}^{n} \log F(z_i)$. Solving $S_M(\alpha) = 0$, we obtain newly

$$\widehat{\alpha}_M = (n-1) \Big/ \left(-\sum_{i=1}^{n} \log F(z_i)\right),$$

as the solution.

## 3. CASE II: $F(\cdot)$ DEPENDS OF $\boldsymbol{\psi}$, A VECTOR OF PARAMETERS

Consider that $F(\cdot)$ is indexed by a vector of parameters $\boldsymbol{\psi}$. In this case, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\psi}, \alpha)$ is given by

$$\ell(\boldsymbol{\theta}) = n \log \alpha + \sum_{i=1}^{n} \Big[\log f(z_i; \boldsymbol{\psi}) + (\alpha - 1) \log F(z_i; \boldsymbol{\psi})\Big].$$

Our proposal is to consider the bias correction only for $\alpha$ and not for $\boldsymbol{\psi}$. This is justifiable in some models such as EE and PHN because the bias for $\alpha$ is considerable in small and median sample sizes and lower for the components of $\boldsymbol{\psi}$ as presented in Figures 1 and 2.
Following the approach used in Sartori [41], we first compute the constrained MLE $\widehat{\boldsymbol{\psi}}(\alpha)$ for fixed $\alpha$, and then we apply Firth's method to the profile score function of $\alpha$, which produces the modified estimator

$$\widehat{\alpha}_M = \frac{n-1}{-\sum_{i=1}^{n} \log F(x_i; \widehat{\boldsymbol{\psi}})}.$$

In short, the estimation procedure can be described as:

- Step 0: Choose an initial value for $\boldsymbol{\theta} = (\boldsymbol{\psi}, \alpha)$, say $\widehat{\boldsymbol{\theta}}^{(0)}$. A possible value can be $\widehat{\boldsymbol{\theta}}^{(0)} = (\widehat{\boldsymbol{\psi}}^{(0)}, 1)$, where $\widehat{\boldsymbol{\psi}}^{(0)}$ is the MLE for $\boldsymbol{\psi}$ considering that $X_1, ..., X_n$ are iid from $F(\cdot; \boldsymbol{\psi})$.

- Step 1: For $k = 1, 2, \ldots$, choose $\widehat{\boldsymbol{\psi}}^{(k)}$ as the vector that maximizes

$$\ell_p(\boldsymbol{\psi}; \widehat{\alpha}^{(k-1)})$$

  in relation to $\boldsymbol{\psi}$.
- Step 2: For $k = 1, 2, \ldots$, do

$$\widehat{\alpha}_M^{(k)} = \frac{n-1}{-\sum\limits_{i=1}^{n} \log F\left(x_i; \widehat{\boldsymbol{\psi}}^{(k)}\right)}.$$

Although we apply the bias correction only for $\alpha$, we will see in the next section that this procedure also provides better estimates for $\boldsymbol{\psi}$.

**Remark 3.1.** Given the MLE of $\boldsymbol{\psi}$, say, $\widehat{\boldsymbol{\psi}}$ and considering Remark 2.1, we can compute the corrective method of Cox–Snell for $\alpha$. This estimator will be denoted as $\widehat{\alpha}_C$ and is given by

$$\widehat{\alpha}_C = \frac{n-1}{-\sum_{i=1}^{n} \log F(z_i; \widehat{\boldsymbol{\psi}})}.$$

Note that in this procedure $\widehat{\boldsymbol{\psi}}$ is not recomputed and matched directly with the MLE estimator. However, to avoid confusion in the simulation study, we consider the notation $\widehat{\boldsymbol{\theta}}_C = (\widehat{\boldsymbol{\psi}}_C, \widehat{\alpha}_C)$ to refer to the estimators obtained by this method.

## 4. SIMULATION STUDY

In this section, we illustrate the method discussed in Section 3 for the EE, PHN and PPE models (see Gómez *et al.* [8]). All the computational programs were developed in R Core Team [39] and are available upon request. Random samples for those distributions can be obtained using the inverse transformation method, considering that the inverse of the cdf for the basal models are implemented in R. We consider sample sizes ranging from 10 to 100, taking one sample for every 5 units. For the EE model, we consider all combinations among the sets $\mathcal{A} = \{0.25, 0.5, 1, 2, 5, 10\}$ and $\mathcal{L} = \{0.1, 0.5, 2\}$ for $\alpha$ and $\lambda$, respectively. In a similar manner, for the PHN model we consider all combinations among the sets $\mathcal{A} = \{0.5, 2, 5, 10\}$ and $\mathcal{S} = \{5, 30, 50\}$ for $\alpha$ and $\sigma$, respectively. For the PPE models we choose a different way to select the parameters. We consider the case $L = 2$, which includes three parameters for the model. For a given time partition $a$ and $\alpha$, we take $\lambda_1 = -(1/a)\log(1 - 0.6^{1/\alpha})$, which guarantees that each observation belongs to the intervals $(0, a)$ and $(a, \infty)$ with probabilities 0.6 and 0.4, respectively. We consider $\alpha$ in $\mathcal{A} = \{0.5, 2, 5, 10\}$, $a$ in $\{6, 10\}$ and $\lambda_2 = 1$ for all combinations.

We consider 10,000 replicates for each combination between $n$, the sample size, $\alpha$ and $\lambda$, $\sigma$ or $(\lambda_1, \lambda_2)$ (depending on the model). In each replication we compute the ordinary MLEs, the estimators considering the Cox–Snell corrective method, and the proposed modified MLEs. For each scenario, we present the relative bias and the relative root mean squared error $\sqrt{MSE}$. In Figure 3 we can find the bias, and in Figure 4 we see the $\sqrt{MSE}$ for one case in the EE distribution. The remaining combinations for the EE, PHN and PPE models are presented as supplementary material.

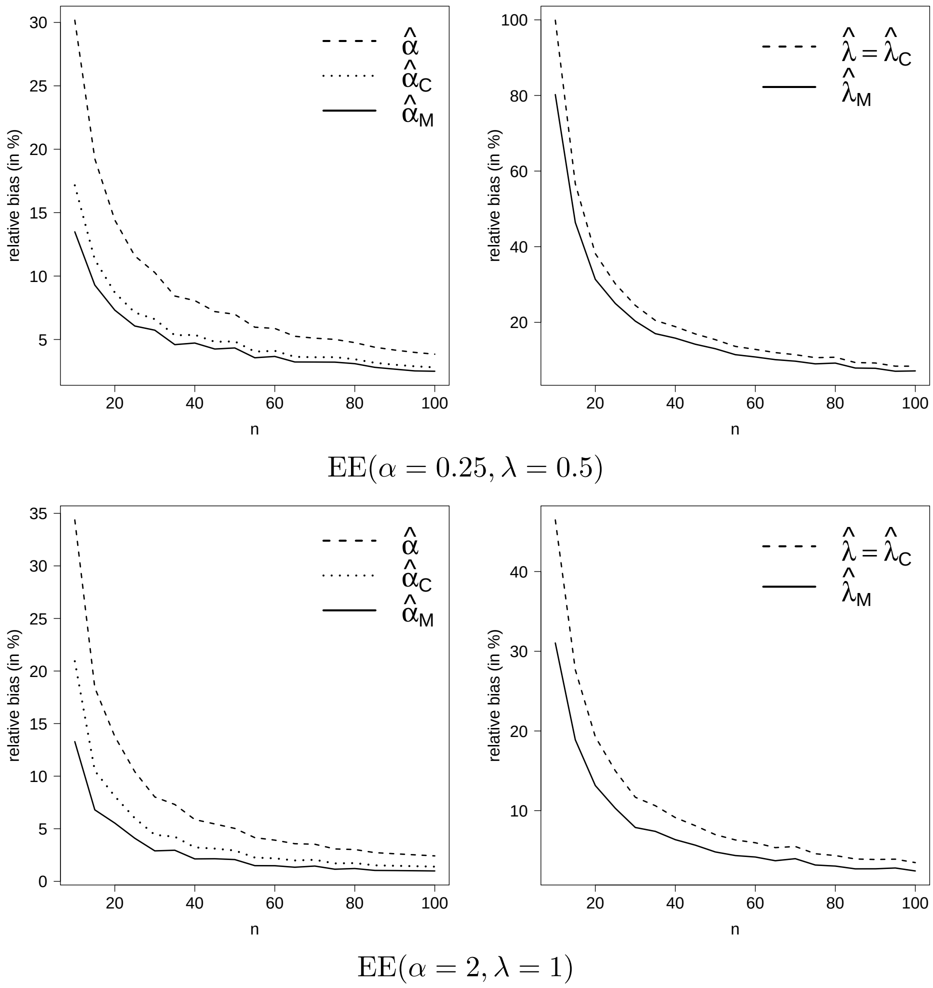
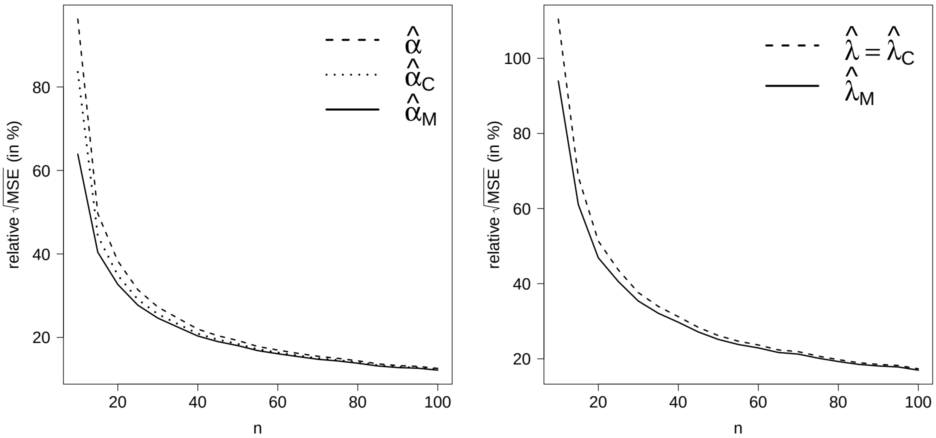



**Figure 3**: Estimated bias for the MLE and the modified MLE in the EE($\alpha, \lambda$) model under different scenarios based on 10,000 replicates.

Note that the bias of $\widehat{\alpha}_M$ is reduced considerably when compared with $\widehat{\alpha}$ and $\widehat{\alpha}_C$ in the three models, EE, PHN and PPE, especially for small sample sizes (say $n \leq 20$). Specifically, for the EE and PHN distributions in all the cases considered for $n = 10$ (the smallest sample size), the bias reduction is at least 10% when $\widehat{\alpha}$ is compared with $\widehat{\alpha}_M$ and at least 5% when is compared $\widehat{\alpha}_C$ with $\widehat{\alpha}_M$. For the PPE distribution in all the cases considered for $n = 10$, the bias reduction is at least 40% when $\widehat{\alpha}$ is compared with $\widehat{\alpha}_M$ and at least 30% when $\widehat{\alpha}_C$ is compared with $\widehat{\alpha}_M$. In all the models, the difference is even greater when the true value of $\alpha$ is increased. On the other hand and as expected, this difference is reduced when the sample size is increased.




EE($\alpha = 2, \lambda = 0.05$)

**Figure 4**: Estimated root MSE for the MLE and the modified MLE in the EE($\alpha = 0.5, \lambda = 0.05$) model under different scenarios based on 10,000 replicates.

The components of the vector $\boldsymbol{\psi}$ also are benefited in terms of bias, although the bias reduction is only proposed for $\alpha$. For the EE model in all the cases considered and $n = 10$, the bias reduction is at least 10% when $\widehat{\lambda}$ or $\widehat{\lambda}_C$ is compared with $\widehat{\lambda}_M$. For the PHN and PPE models, in all the cases considered the bias reduction for $\lambda$ and $(\lambda_1, \lambda_2)$ exists, but is marginal.

Additionally, $\sqrt{MSE}$ related to $\widehat{\alpha}_M$ is also lower when compared with $\widehat{\alpha}$ and $\widehat{\alpha}_C$. For the EE and PHN distributions, in all the cases considered for $n = 10$ the $\sqrt{MSE}$ is reduced by at least around 5% when $\widehat{\alpha}$ or $\widehat{\alpha}_C$ is compared with $\widehat{\alpha}_M$, whereas for the PPE model in all the cases considered for $n = 10$ this reduction is at least around 20%. On the other hand, for the EE and PPE models in all the cases considered for $n = 10$, the $\sqrt{MSE}$ is reduced by at least around 10% when the modified estimator is compared with the traditional estimator or the Cox–Snell estimator; whereas for the PHN distribution, in all the considered models the reduction for $\sqrt{MSE}$ is marginal. Again, in all the models, the difference is even greater when the true value for $\alpha$ is increased and as expected, the difference between the different estimators is reduced when the sample size is increased.

These simulation results are encouraging since they show, for these three particular members of the class of power models, that even though we focus this bias prevention method on the shape parameter, $\alpha$, we still observe better bias results for the other parameters.

## 5. APPLICATIONS

In this section, we illustrate the methods in three real data sets for the EE, PPE and PN models. All data sets are already known in the literature and we wish to compare the performance of the modified MLE against the ordinary MLE and also using the Cox–Snell correction method. We examine not only how well they both fit the data, but also the bias, which could be estimated through bootstrap. Additional applications for the PHN and PBS models are presented as supplementary material.

## 5.1. Illustration 1

In this first application, we consider data on the number of million revolutions before failure for each of 23 ball bearings in a life test. More details about the data are presented in Lawless [17]. This data set was analyzed in Gupta and Kundu [12] using the EE model. The estimates considering the ordinary MLE's and the modified MLEs for this model are presented in Table 2, with respective s.e. and confidence intervals.

**Table 2**: MLE and modified MLE for the EE model in the ball bearings data set.

| Parameter | $\widehat{\lambda}$ | $\widehat{\alpha}$ | $\widehat{\lambda}_M$ | $\widehat{\alpha}_M$ |
|---|---|---|---|---|
| Estimate | 0.0323 | 5.2832 | 0.0302 | 4.5379 |
| s.e. | 0.0064 | 2.0492 | 0.0060 | 1.6566 |
| 95% C.I. | (0.0197 ; 0.0449) | (1.2667 ; 9.2998) | (0.0206 ; 0.0440) | (2.0363 ; 8.5302) |

Note that the confidence intervals when we consider the modified MLEs are more accurate for both parameters, since they have a smaller length compared to the confidence intervals obtained with the estimates of the ordinary MLEs. The histogram of the data and the estimated density functions of both estimates are presented in Figure 5a.

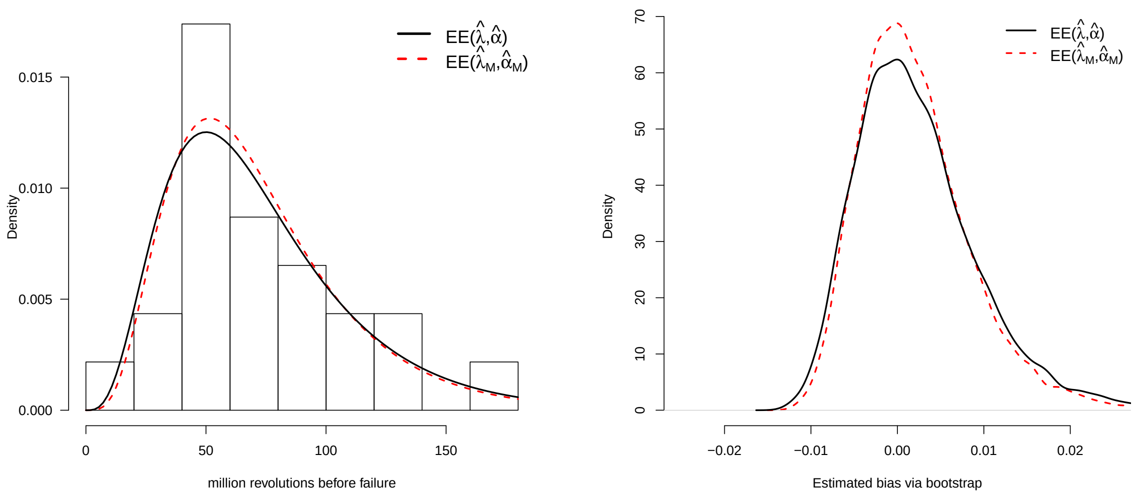



**Figure 5**: Comparisons for the ordinary and modified MLEs in the ball bearings data set. Left panel: Histogram for EE model and respective density estimate for each proposal. Right panel: Estimated bias via 10,000 bootstrap samples for each estimator.

The estimated density for the modified MLEs presents a better fit with the data, if we note that this density better represents the peak of values around 50 million revolutions before failure. Furthermore, if we take 10,000 bootstrap samples to estimate the bias, we have Figure 5b, where we are able to compare these two estimators empirically. The ordinary MLE present an estimated bias, via bootstrap, equal to 1.4774, while this value for the modified bias is 1.0626. The bootstrap standard errors are equal to 3.8744 and 2.7440 for the ordinary and modified MLE, respectively. These results are confirmed by the analysis of Figure 5b, where the estimated density for the modified MLE is clearly more concentrated around zero, further evidence of the superior performance of the modified method.

## 5.2. Illustration 2

This data set is available in Murthy *et al.* [34] (data set 6.1 in Section 6.6.5). The data set represents the failure time of 20 components. We propose to analyze this data set based on the PPE model with $L = 2$ in order to illustrate the advantage of our methodology to reduce the bias for parameters. Table 3 shows the ordinary, corrected and modified MLEs for the PPE distribution in this data set. The main differences between the three methods are given in the estimates for $\lambda_1$ and $\alpha$. We also highlight that the standard errors are lower for the modified MLEs, which also provides more accurate confidence intervals.

**Table 3**: Ordinary, corrected and modified MLE for the PPE model in failure time data set.

| Parameter | Estimate | s.e. | 95% C.I. |
|---|---|---|---|
| $\widehat{\lambda}_1$ | 0.8766 | 0.2643 | (0.4855 ; 1.5829) |
| $\widehat{\lambda}_2$ | 3.8663 | 1.1347 | (2.1751 ; 6.8725) |
| $\widehat{\alpha}$ | 5.1751 | 2.6060 | (1.9287 ; 13.8856) |
| $\widehat{\lambda}_{1C}$ | 0.8766 | 0.2477 | (0.4967 ; 1.5472) |
| $\widehat{\lambda}_{2C}$ | 3.8663 | 1.1394 | (2.1759 ; 6.8699) |
| $\widehat{\alpha}_C$ | 4.9163 | 2.2601 | (1.8781 ; 12.8693) |
| $\widehat{\lambda}_{1M}$ | 0.7600 | 0.2366 | (0.3893 ; 1.4836) |
| $\widehat{\lambda}_{2M}$ | 3.8379 | 1.1303 | (2.1542 ; 6.8376) |
| $\widehat{\alpha}_M$ | 4.0096 | 1.8634 | (1.4397 ; 11.1671) |

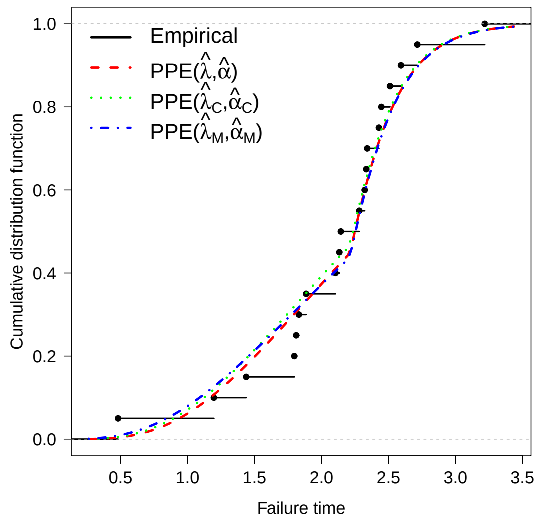



**Figure 6**: Cumulative distribution function for failure time data set using PPE model for ordinary MLE, corrected MLE and modified MLE.

Figure 6 shows the empirical cdf and the estimated cdf for the three methods. The main difference between the curves is given before the median of the distribution (approximately 2.2 units). Finally, Figure 7 shows the estimated distribution for the bias of the estimators of $\lambda_1$, $\lambda_2$ and $\alpha$ for the three methods, which are computed based on 10,000 bootstrap samples. Again, the main differences are given for the estimators for $\lambda_1$ and $\alpha$. For this last term, the estimators provided by the ordinary and corrected MLEs have an evident and considerable bias, in contrast to the modified MLE where the bias is negligible.

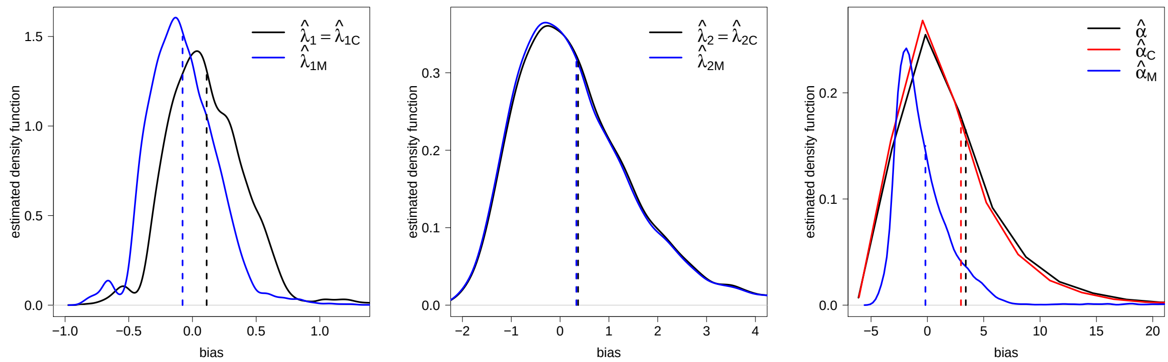



**Figure 7**: Estimated distribution for the bias for MLEs, corrected MLEs and modified MLEs based on 10,000 bootstrap samples. Dashed lines represents the respective average bias.

## 5.3. Illustration 3

This data set is related to 3,848 observations of the variable "density" in the data available at `http://lib.stat.cmu.edu/datasets/pollen.data` and was analyzed by Pewsey *et al.* [38] using the PN distribution. Although the PN distribution was not considered in the simulation studies, we decided to include this example using the PN model in order to demonstrate the effectiveness of our modified MLE for the parameters in a class of location-scale ($\mu$ and $\sigma$, respectively) within the power models. Evidently, as the sample size is large, it is to be expected that the ordinary MLE and their corrections will be closer. However, in order to illustrate our proposal, we considered a subsample of $n = 30$ from the original data. Table 4 presents the results. Note that the estimates for $\alpha$ are closer for the ordinary MLE and the Cox–Snell corrective method, but differ strongly from the modified MLE. The impact of this can be assessed by the huge reduction in the bias of $\widehat{\alpha}_M$ in comparison with the bias of $\widehat{\alpha}$ and $\widehat{\alpha}_C$. In addition, Figure 8 shows the histogram with the estimated density function based on the PN model for the three estimation methods.

**Table 4**: Ordinary, corrected, modified MLE and estimated bias based on 1,000 non-parametric bootstrap samples for the PN model in the pollen data set.

| Parameter | Estimate | s.e. | 95% C.I. | bias |
|---|---|---|---|---|
| $\widehat{\mu}$ | −16.3238 | 36.7854 | (−53.1092 ; 20.4616) | 10.6265 |
| $\widehat{\sigma}$ | 6.8207 | 6.5279 | (0.2928 ; 13.3486) | −2.8944 |
| $\widehat{\alpha}$ | 115.1611 | 970.8617 | (0.0000 ; 1086.0228) | 265.4226 |
| $\widehat{\mu}_C$ | −16.3238 | 5.3937 | (−21.7175 ; −10.9301) | 10.6265 |
| $\widehat{\sigma}_C$ | 6.8207 | 1.3255 | (5.4952 ; 8.1462) | −2.8944 |
| $\widehat{\alpha}_C$ | 111.3105 | 117.7705 | (0.0000 ; 229.0813) | 256.5461 |
| $\widehat{\mu}_M$ | 4.7456 | 1.6875 | (3.0581 ; 6.4331) | 1.1611 |
| $\widehat{\sigma}_M$ | 1.7058 | 0.5620 | (1.1438 ; 2.2678) | −0.4343 |
| $\widehat{\alpha}_M$ | 0.1838 | 0.1674 | (0.0163 ; 0.3512) | −0.0601 |

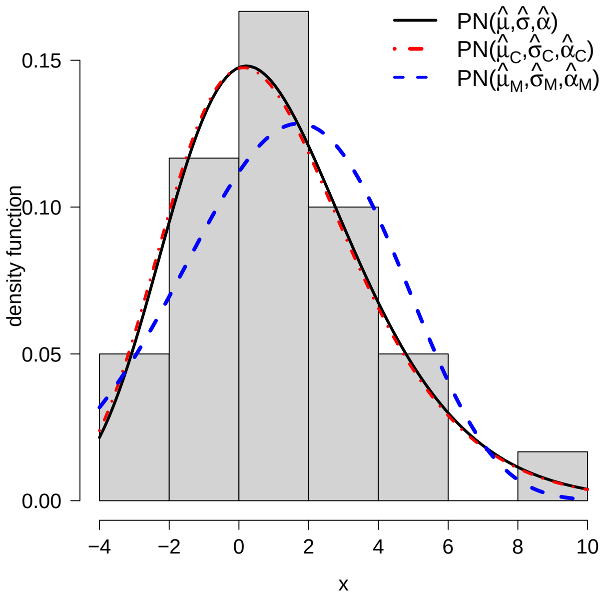



**Figure 8**: Estimated pdf for pollen data set using PN model for ordinary MLE, corrected MLE and modified MLE.

## 6. CONCLUSIONS

In this paper, we considered the problem of estimation of the shape parameter in the general class of power models; we recognize empirically the difficulties in this task, especially for small sample sizes. For the case where we have only $\alpha$ to estimate, we present the unbiased estimator as a function of a complete statistic for this class of models, obtaining the UMVUE for $\alpha$. We discuss the connection of these results with the bias prevention method proposed by Firth [6]. We also propose an estimation method for the case when there are more parameters, limiting the bias correction to $\alpha$.

Although our results are valid for all members of the general class of power models, we selected some members of this class of models, namely EE, PN, PPE, PHN and PBS, in order to demonstrate and compare the results between the ordinary MLE and the modified MLE proposed in this paper. The simulation studies confirm the bias reduction for the shape parameter, but they also show that there is an improvement related to bias for the other parameters involved, for each distribution considered here. According to our simulation results, the improvements are not only related to bias, as we also noticed lower root mean squared errors when we use the modified estimator. Although we do not consider families of bimodal distributions belonging to this family (as presented in Bolfarine *et al.* [4]), we see no reason why the method should not work in families of this type.

We illustrate our findings with three known data sets from the literature, for each distribution. Although this is a large number of examples, we thought it was important to make sure our method was tested with different members of this class, and not just some selected cases. We show that our modified estimator gives a better fit with the data in each case, and also we estimate the bias via bootstrap, validating that our proposal performs better.

## ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees for their comments and/or suggestions, which undoubtedly improved the presentation of our article.

The research of O. Venegas was supported by Vicerrectoría de Investigación y Postgrado de la Universidad Católica de Temuco, Projecto interno FEQUIP 2019-INRN-03. Héctor W. Gómez's research has been partially supported by PUENTE UA.

## REFERENCES

[1] Arrué, J.; Arellano-Valle, R.B. and Gómez, H.W. (2016). Bias reduction of maximum likelihood estimates for a modified skew normal distribution, *Journal of Statistical Computation and Simulation*, **86**, 2967–2984.

[2] Birnbaum, Z.W. and Saunders, S.C. (1969a). A new family of fife distributions, *Journal of Applied Probability*, **6**, 319–327.

[3] Birnbaum, Z.W. and Saunders, S.C. (1969b). Estimation for a family of life distributions with applications to fatigue, *Journal of Applied Probability*, **6**, 328–347.

[4] Bolfarine, H.; Martínez-Flórez, G. and Salinas, H.S. (2018). Bimodal symmetric-asymmetric power-normal families, *Communications in Statistics – Theory and Methods*, **47**, 259–276.

[5] Durrans, S.R. (1992). Distributions of fractional order statistics in hydrology, *Water Resources Research*, **28**, 1649–1655.

[6] Firth, D. (1993). Bias reduction of maximum likelihood estimates, *Biometrika*, **80**, 27–38.

[7] Gómez, Y.M. and Bolfarine, H. (2015). Likelihood-based inference for the power half-normal distribution, *Journal of Statistical Theory and Applications*, **14**, 383–398.

[8] Gómez, Y.M.; Gallardo, D.I. and Arnold, B.C. (2017). The power piecewise exponential model, *Journal of Statistical Computation and Simulation*, **88**, 825–840.

[9] Gupta, R.C.; Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data by Lehmann alternatives, *Communications in Statistics – Theory and Methods*, **27**, 887–904.

[10] Gupta, R.D. and Gupta, R.C. (2008). Analyzing skewed data by power normal model, *Test*, **17**, 197–210.

[11] Gupta, R.D. and Kundu, D. (1999). Generalized exponential distributions, *Australian & New Zealand Journal of Statistics*, **41**, 173–188.

[12] Gupta, R.D. and Kundu, D. (2001). Exponentiated exponential family: An alternative to Gamma and Weibull distributions, *Biometrical Journal*, **43**, 117–130.

[13] Ioannis, K.; Pagui, E.C.K. and Sartori, N. (2020). Mean and median bias reduction in generalized linear models, *Statistics and Computing*, **30**, 43–59.

[14] Kakde, C.S. and Shirle, D.T. (2006). On exponentiated lognormal distribution, *International Journal of Agricultural and Statistical Sciencies*, **2**, 319–326.

[15] Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models, *Biometrika,* **96**, 793–804.

[16] Kyriakou, S.; Ioannis, K. and Sartori, N. (2019). Median bias reduction in random-effects meta-analysis and meta-regression, *Statistical Methods in Medical Research*, **28**, 1622–1636.

[17] Lawless, E.L. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.

[18] Lehmann, E.L. (1953). The power of rank tests, *Annals of Mathematical Statistics*, **24**, 23–43.

[19] Lemonte, A. (2007) Improved statistical inference for the two-parameter Birnbaum-Saunders distribution, *Computational Statistics & Data Analysis*, **51**, 4656–4681.

[20] Magalhães, T.M.; Gallardo, D.I. and Bourguignon, M. (2021). Improved point estimation for inverse gamma regression models, *Journal of Statistical Computation and Simulation*, **91**, 2444–2456.

[21] Magalhães, T.M.; Gómez, Y.M.; Gallardo, D.I. and Venegas, O. (2020). Bias reduction for the Marshall-Olkin extended amily of distributions with application to an airplane's air aonditioning aystem and arecipitation data, *Symmetry*, **12**(5), 851.

[22] Maity, A.K.; Pradhan, V. and Das, U. (2019). Bias reduction in logistic regression with missing responses when the missing data mechanism is nonignorable, *The American Statistician*, **73**, 340–349.

[23] Martínez-Flórez, G.; Arnold, B.C.; Bolfarine, H. and Gómez, H.W. (2013a). The multivariate alpha-power model, *Journal of Statiscal Planning and Inference*, **143**, 1236–1247.

[24] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2013b). Asymmetric regression models with limited responses with an application to antibody response to vaccine, *Biometrical Journal*, **55**, 156–172.

[25] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2013c). Skew-normal alpha-power model, *Statistics: A Journal of Theoretical and Applied Statistics*, **48**, 1414–1428.

[26] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2014a). An alpha-power extension for the Birnbaum-Saunders distribution, *Statistics: A Journal of Theoretical and Applied Statistics*, **48**, 896–912.

[27] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2014b). The log alpha-power asymmetric distribution with application to air pollution, *Environmetrics*, **25**, 44–56.

[28] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2015). Doubly censored power-normal regression models with inflation, *Test*, **24**, 265–286.

[29] Martínez-Flórez, G.; Bolfarine, H. and Gómez, H.W. (2017). The log-linear Birnbaum-Saunders power model, *Methodology and Computing in Applied Probability*, **19**, 913–933.

[30] Mazucheli, J.; Menezes, A.F.B.; Alqallaf, F. and Ghitany, M.E. (2021). Bias-corrected Maximum Likelihood estimators of the parameters of the Unit-Weibull distribution, *Austrian Journal of Statistics*, **50**, 41–53.

[31] Melo, T.F.N.; Ferrari, S.L.P. and Patriota, A.G. (2018). Improved estimation in a general multivariate elliptical model, *Brazilian Journal of Probability and Statistics*, **32**, 44–68.

[32] Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate date, *IEEE Transactions on Reliability*, **42**, 299–302.

[33] Mudholkar, G.S.; Srivastava, D.K. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus motor failure data, *Technometrics*, **37**, 436–445.

[34] Murthy, D.P.; Xie, M. and Jiang, R. (2004). *Weibull Models*, John Wiley & Sons, New York.

[35] Nadarajah, S. (2005). The exponentiated Gumbel distribution with climate application, *Environmetrics*, **17**, 13–23.

[36] Nadarajah, S. and Gupta, A.K. (2007). The exponentiated gamma distribution with application to drought data, *Calcutta Statistical Association Bulletin*, **59**, 29–54.

[37] Pagui, K.; Salvan, E.C. and Sartori, N. (2017). Median bias reduction of maximum likelihood estimates, *Biometrika*, **104**, 923–938.

[38] Pewsey, A.; Gómez, H.W. and Bolfarine, H. (2012). Likelihood-based inference for power distributions, *Test*, **21**, 775–789.

[39] R Core Team. (2021). *R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing*, ISBN: 3-900051-07-0, version 4.1.1

[40] Rieck, J.R. and Nedelman, J.R. (1991). A log-linear model for the Birnbaum-Saunders distribution, *Technometrics*, **33**, 51–60.

[41] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions, *Journal of Statistical Planning and Inference*, **136**, 4259–4275.

[42] Zhao, J. and Kim, H.M. (2016). Power t distribution, *Communications for Statistical Applications and Method*, **23**, 321–334.