






Article

A New Wavelet-Based Privatization Mechanism for Probability Distributions

Hélio M. de Oliveira ¹, Raydonal Ospina ¹, Víctor Leiva ^{2,*}, Carlos Martin-Barreiro ^{3,4}
and Christophe Chesneau ⁵

¹ Department of Statistics, CASTLab, Universidade Federal de Pernambuco, Recife 50670-901, Brazil; hmo@de.ufpe.br (H.M.d.O.); raydonal@de.ufpe.br (R.O.)

² School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile

³ Faculty of Natural Sciences and Mathematics, Escuela Superior Politécnica del Litoral ESPOL, Guayaquil 090902, Ecuador; cmmartin@espol.edu.ec

⁴ Faculty of Engineering, Universidad Espíritu Santo, Samborondón 0901952, Ecuador

⁵ Department of Mathematics, Université de Caen Basse-Normandie, F-14032 Caen, France; christophe.chesneau@unicaen.fr

* Correspondence: victor.leiva@pucv.cl or victorleivasanchez@gmail.com

Abstract: In this paper, we propose a new privatization mechanism based on a naive theory of a perturbation on a probability using wavelets, such as a noise perturbs the signal of a digital image sensor. Wavelets are employed to extract information from a wide range of types of data, including audio signals and images often related to sensors, as unstructured data. Specifically, the cumulative wavelet integral function is defined to build the perturbation on a probability with the help of this function. We show that an arbitrary distribution function additively perturbed is still a distribution function, which can be seen as a privatized distribution, with the privatization mechanism being a wavelet function. Thus, we offer a mathematical method for choosing a suitable probability distribution for data by starting from some guessed initial distribution. Examples of the proposed method are discussed. Computational experiments were carried out using a database-sensor and two related algorithms. Several knowledge areas can benefit from the new approach proposed in this investigation. The areas of artificial intelligence, machine learning, and deep learning constantly need techniques for data fitting, whose areas are closely related to sensors. Therefore, we believe that the proposed privatization mechanism is an important contribution to increasing the spectrum of existing techniques.

Keywords: artificial intelligence; data fitting; database-sensor; digital image sensor; machine learning; perturbation theory; signal-to-noise ratio; statistical modeling; wavelets



Citation: de Oliveira, H.M.; Ospina, R.; Leiva, V.; Martin-Barreiro, C.; Chesneau, C. A New Wavelet-Based Privatization Mechanism for Probability Distributions. *Sensors* **2022**, *22*, 3743. <https://doi.org/10.3390/s22103743>

Academic Editors: Weiming Shen, Giancarlo Fortino, Antonio Liotta, Tie Qiu, Wenfeng Li and Yanjun Shi

Received: 20 April 2022

Accepted: 11 May 2022

Published: 14 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Probability models capable of capturing the fundamental information contained in modern data, as those used for artificial intelligence [1] and big data [2], as well as models presenting unique features, have promoted derivations of novel continuous probability distributions [3,4].

Numerous and diverse approaches have been proposed over time to generate new probability or statistical distributions [5]. One of the most common approaches allows us to enhance the functionality of a base continuous cumulative distribution function (CDF). This can be achieved utilizing various transformations based on exponential, logarithmic, power, or other functions [6].

On this topic, we may refer to the so-called “families of probability distributions”, as described in [7,8]. The new probability distributions may be employed efficiently in diverse settings, as described in [9,10]. We may also refer to the work stated in [11] pointing out the importance of continuous probability distributions in the definition of various measures.

In view of the impacts of the current research on probability distributions [12], diverse applications related to the areas of artificial intelligence [1], machine learning [13], and deep learning [14] constantly require new techniques for data fitting, whose areas are closely related to sensors. Additionally, to aid in the progress of computer sciences, new approaches are welcome to expand the options of a reference probability distribution [15].

An application of probability models can be introduced by perturbing a CDF additively, similarly to how a noise perturbs the signal of a digital image sensor [16]. Surprisingly, such a strategy does not appear to have received much attention in the literature. More precisely, given a continuous CDF, one can add this function to another (the perturbation function) in such a way that the resulting function is also a continuous CDF.

To propose a manageable perturbation [17], one can employ a special, well-known function called wavelet [18,19]. Basically, such a function has a wave-like oscillation with an amplitude that starts at zero and increases or decreases before returning to zero, one or more times. Wavelets may be utilized to extract information from a wide range of data, including audio signals and images often related to sensors [20], as unstructured data. To thoroughly analyze data, wavelet sets might be used. For more information on wavelets, we refer the reader to [21–23]. More specifically, in [24], transients and their wavelet coefficients are modeled as mixed Laplace probability density functions (PDFs). In [25], image segmentation based on a wavelet feature descriptor and dimensionality reduction was applied to remote sensing. Thus, one could involve a wavelet function to define a valid perturbation, and then a privatized probability distribution can be obtained through theoretical and practical tools.

The main objectives of this article are to propose and derive a naive theory of an additive perturbation on a continuous probability distribution based on a wavelet approach, and to illustrate it with a sensor-related application. The use of wavelets in this probability distribution setting is original, and our findings offer up a new modeling horizon, which are examined in depth. Therefore, we offer a mathematical method for choosing a suitable probability distribution to model data by starting from some guessed-at initial probability distribution. Examples for the proposed method are also presented. For the computational experiments, we utilize a database-sensor and two related algorithms.

The rest of the article is organized as follows. Section 2 introduces the new wavelet approach. In Section 3, we discuss the choice of a perturbation for an arbitrary probability distribution. Section 4 proposes a correction for statistical moments due to the perturbation. Then, in Section 5, the generalization of the perturbation approach at further levels is presented. In Section 6, we provide an empirical application of our approach. Finally, Section 7 gives the concluding remarks.

2. Background and Wavelet Approach

Suppose we have a random variable X with a continuous CDF F_X . Let us consider an additive (functional) perturbation, denoted as ε -perturbation, so that

$$F_{\text{priv}}(x) := F_X(x) + \varepsilon(x), \quad (1)$$

with the CDF F_{priv} stated in (1) being a privatized CDF.

Note that, in the expression defined in (1), the CDF of the variable X has been perturbed and a new function F_{priv} is obtained. However, the choice of the perturbation cannot be arbitrary because it could break the requirements to deal only with a probability distribution. The following conditions must be met by the perturbation:

(C1) $\lim_{|x| \rightarrow +\infty} \varepsilon(x) = 0$;

(C2) ε is derivable and satisfies $|\mathrm{d}\varepsilon(x)/\mathrm{d}x| \leq f_X(x)$, where f_X denotes the PDF related to the CDF F_X .

The conditions (C1) and (C2) above stated guarantee that F_{priv} is also a CDF. This new distribution could be seen as a privatized version of the reference distribution.

To describe our new wavelet approach, some definitions need to be given. Let us begin with the mathematical definition of a wavelet.

Definition 1 (Wavelet function). *A wavelet is a Lebesgue measurable function $\psi(x)$ that is both absolutely integrable and square-integrable, such that*

$$\int_{-\infty}^{+\infty} \psi(x) dx = 0, \quad (2)$$

$$\int_{-\infty}^{+\infty} \psi^2(x) dx = 1. \quad (3)$$

On the one hand, from the expression established in (2), observe that the absolute value of ψ is integrable over the entire real line and its result is equal to zero (0). On the other hand, in the formula stated in (3), note that the square of ψ is also integrable over \mathbb{R} and its result is equal to one (1). Keep in mind that, in this study, we deal with compactly supported wavelets [26], that is, the closure of the set upon which the wavelet stands non-vanishing is a compact set. Specifically, if ψ is a wavelet function, then $\{x: \psi(x) \neq 0\}$ is a compact set, and we say ψ is a wavelet of compact support. Henceforth, we assume that $\text{support}\{\psi(x)\} \equiv [a, b]$, which plays a crucial role in our proposal [21,27]. The next definition presents the notion of wavelet cumulative function in this setting.

Definition 2 (Wavelet cumulative function). *A wavelet cumulative function is defined by*

$$\Psi(x) := \int_{-\infty}^x \psi(\zeta) d\zeta. \quad (4)$$

Since only compactly supported wavelets are considered, the wavelet cumulative function given in (4) can be simplified to

$$\Psi(x) = \int_a^x \psi(\zeta) d\zeta, \quad a \leq x \leq b. \quad (5)$$

Thus, from the expression stated in (5), the following properties can be verified:

$$\Psi(x) = \Psi(a) = 0, \quad x \leq a, \quad (6)$$

$$\Psi(x) = \Psi(b) = 1, \quad x \geq b, \quad (7)$$

$$\frac{d\Psi(x)}{dx} = \psi(x). \quad (8)$$

Note that the properties formulated in (6)–(8) are helpful. To begin with, let us deal with the uniform distribution, denoted as $\mathcal{U}[0, 1]$, whose CDF is given by $F_X(x) = x$, for $0 \leq x \leq 1$, where $F_X(x) = 0$, for $x \leq 0$, and $F_X(x) = 1$, for $x \geq 1$. A mapping is proposed to bring the support $[0, 1]$ of the uniform distribution to the support $[a, b]$ of the wavelet, that is, $[0, 1] \xrightarrow{\text{map}} [a, b]$. Then, we propose to choose a particular perturbation ε according to

$$\varepsilon(x) := \Psi_{[0,1]}(x) = \frac{1}{(b-a)} \Psi((b-a)x + a). \quad (9)$$

For the particular choice stated in (9), the new distribution defined in (1) has the same support as the original distribution, with no perturbation added. Furthermore, imposing the condition $|\psi(t)| \leq 1$, it follows that

$$|\varepsilon(x)| \leq \frac{1}{(b-a)} \int_a^{(b-a)x+a} |\psi(\zeta)| d\zeta. \quad (10)$$

From the expression established in (10), we can guarantee that $|\varepsilon(x)| \leq x$, for all $x \in [0, 1]$. Therefore, the condition $F_{\text{priv}}(x) \geq 0$ is assured, for all $x \in [0, 1]$. Hence, we must determine whether F_{priv} is always a non-descending function or not. Thus, we examine the behavior of the corresponding PDF formulated as

$$f_{\text{priv}}(x) = \frac{dF_{\text{priv}}(x)}{dx} = 1 + \frac{1}{(b-a)} \frac{d\Psi((b-a)x+a)}{dx}, \quad (11)$$

implying

$$f_{\text{priv}}(x) = 1 + \psi((b-a)x+a), \quad (12)$$

where f_{priv} denotes the PDF related to the CDF F_{priv} .

From the formulas given in (11) and (12), it follows that $\int_{-\infty}^{+\infty} f_{\text{priv}}(x)dx = 1$ and $f_{\text{priv}}(x) \geq 0$, for all x , thereby proving that this is indeed a valid PDF to be considered. Then, this new PDF and its associated CDF might be visualized as a privatized version of the reference distribution, with the privatization mechanism being named wavelet perturbation. This is that we call “privatization analysis”.

As an example, let us first consider a compactly supported wavelet defined within $[0, 1]$ proposed in [28] and mathematically defined as

$$\psi_{\mathcal{U}}(x) := -\frac{1}{2}x \ln(x) + \frac{1}{2}(1-x) \ln(1-x). \quad (13)$$

Figure 1 shows the original distribution, that is, $\mathcal{U}[0, 1]$, and the new distribution generated by the perturbation identified in (13).

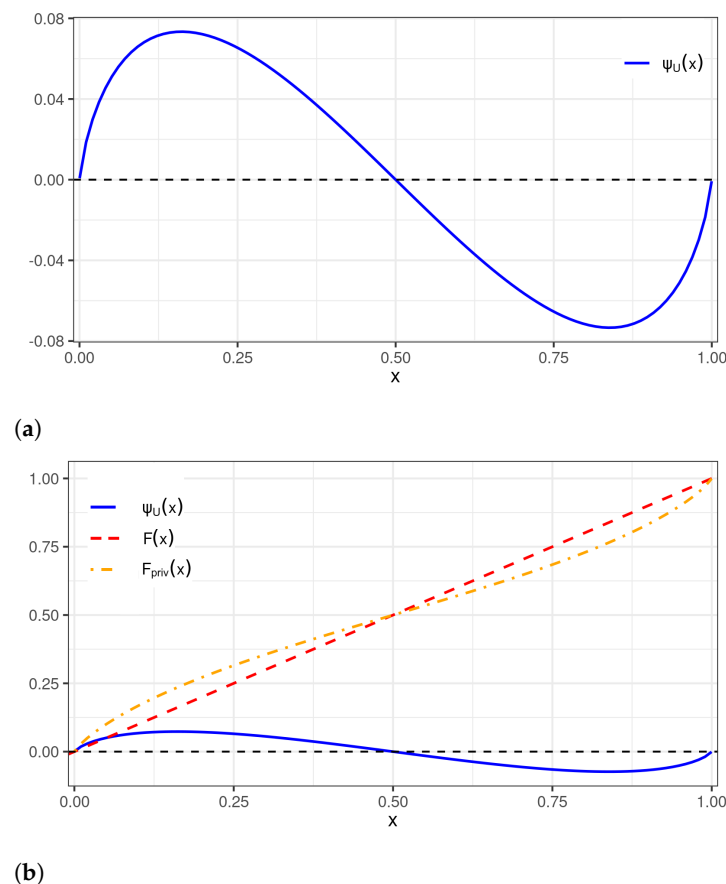


Figure 1. Plots of: (a) a wavelet perturbation to be applied to the $\mathcal{U}[0, 1]$ distribution; and (b) wavelet perturbation (— blue), uniform (- - red), and perturbed uniform (- · - orange) CDFs.

Another family of compactly supported wavelets with parameters that can be adjusted is the beta wavelet family [29]. One of the advantages of adopting beta wavelet perturbations consists of the easy replacement of shape ($\alpha > 0$) and scale ($\theta > 0$) parameters to make the perturbation $\psi_{\text{beta}}(x, \alpha, \theta)$ flexible. In other words, this wavelet family allows for a simple parametrization that drives the asymmetry of the resulting probability distribution. The plots of two beta wavelet perturbations are shown in Figure 2 as examples.

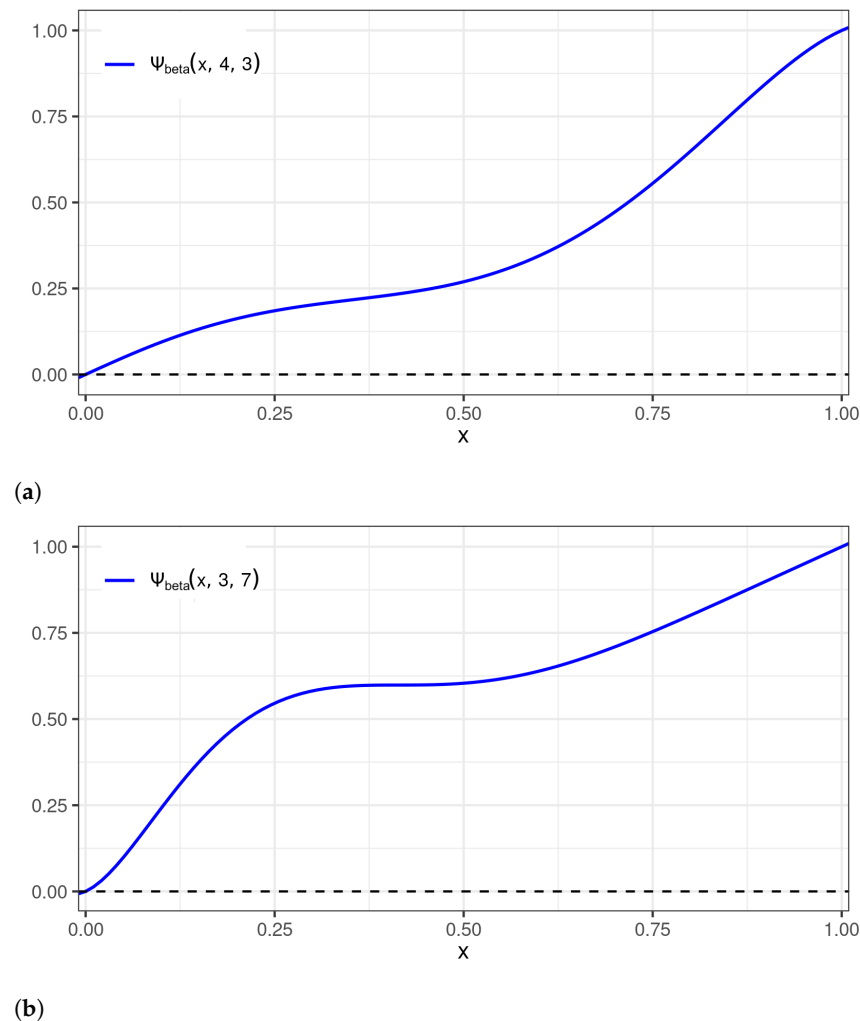
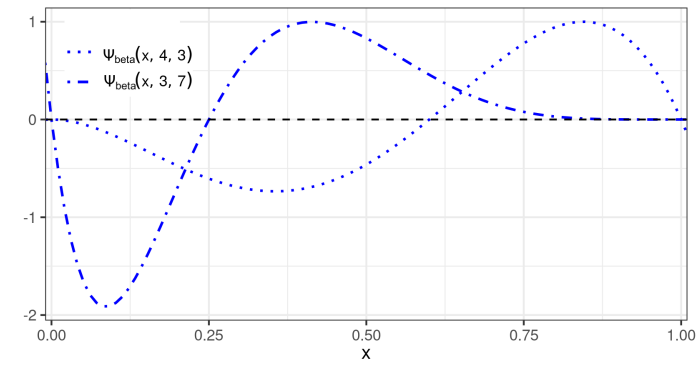
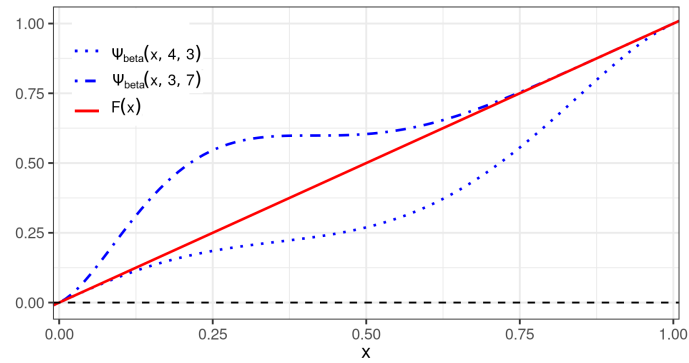


Figure 2. Plots of the beta wavelet perturbations: (a) $\psi_{\text{beta}}(x, 4, 3)$; and (b) $\psi_{\text{beta}}(x, 3, 7)$.

Figure 3 displays perturbed uniform distributions that are generated as a result of applying the perturbations of Figure 2. This approach can be employed to introduce asymmetries in a chosen probability distribution, controlled by the beta wavelet parameter. Among the compactly supported wavelets, certainly the most used are the Daubechies (DB4) wavelets [27]. Expressions close to approximately the DB4 wavelets of any order have been proposed in [30]. Using Matlab™ commands, these continuous approximations were employed to plot the DB4 perturbation adapted to the $\mathcal{U}[0, 1]$ distribution, denoted by Ψ_{DB4} , in Figure 4.

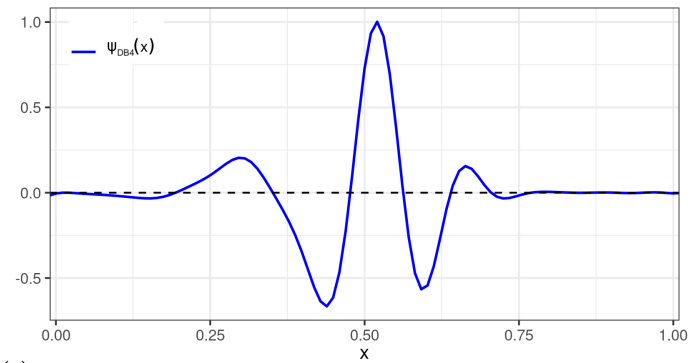


(a)

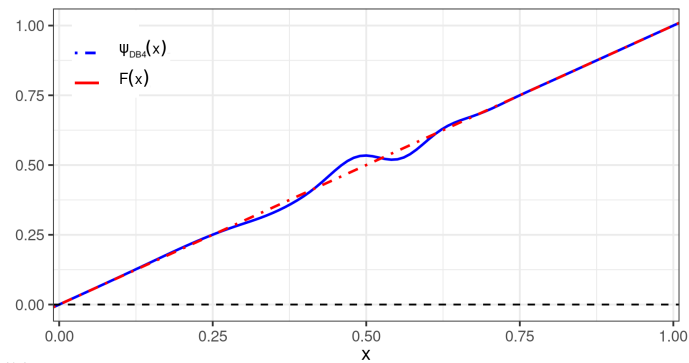


(b)

Figure 3. Plots of: (a) beta wavelet perturbations to be applied to the $\mathcal{U}[0,1]$ distribution; and (b) $\psi_{\text{beta}}(x,4,3)$ perturbed uniform (\cdots blue), $\psi_{\text{beta}}(x,3,7)$ perturbed uniform ($-\cdots$ blue), and uniform ($—$ red) CDFs.



(a)



(b)

Figure 4. Plots of: (a) a DB4 wavelet perturbation to be applied to the $\mathcal{U}[0,1]$ distribution; and (b) DB4 wavelet perturbation ($—$ blue) and uniform ($-\cdots$ red) CDFs.

3. Choosing a Perturbation for an Arbitrary Probability Distribution

Now, we offer a valid perturbation for an arbitrary CDF F_X . For a given compactly supported wavelet ψ with its cumulative function (see Definition 2), consider a new chosen CDF according to

$$F_{\text{priv}}(x) := F_X(x) + \varepsilon(x), \quad (14)$$

with

$$\varepsilon(x) := \frac{1}{(b-a)} \frac{\Psi((b-a)F_X(x) + a)}{\max_{\zeta \in [a,b]} |\psi(\zeta)|}.$$

From (11) and (14), note that $F_{\text{priv}}(-\infty) = 0$, $F_{\text{priv}}(+\infty) = 1$, and

$$f_{\text{priv}}(x) = \frac{dF_{\text{priv}}(x)}{dx} = f_X(x) + \frac{d\varepsilon(x)}{dx}, \quad (15)$$

with $d\varepsilon(x)/dx$ stated in (15) given by

$$\frac{d\varepsilon(x)}{dx} := \frac{\psi((b-a)F_X(x) + a)}{\max_{\zeta \in [a,b]} |\psi(\zeta)|} f_X(x). \quad (16)$$

Then, ε is a valid perturbation because the condition (C1) is satisfied. In addition, we have $\lim_{|x| \rightarrow +\infty} \varepsilon(x) = 0$ due to $\int_a^b \psi(u) du = 0$, so that the condition (C2) is also satisfied, since

$$\left| \frac{\psi((b-a)F_X(x) + a)}{\max_{\zeta \in [a,b]} |\psi(\zeta)|} \right| \leq 1, \quad (17)$$

by (16), having $|d\varepsilon(x)/dx| \leq f_X(x)$. Thus, any wavelet of compact support can be used to induce a different perturbation in the vicinity of the probability distribution initially assigned. From the expressions stated in (14)–(17), note that, after applying the perturbation, the resulting function is also a CDF.

In summary, given a random variable X with CDF F_X , a perturbation can be added, which guarantees that the modified function is still a CDF around the original CDF. This new CDF, and its associated distribution, as mentioned, are privatized versions of the reference distribution using a wavelet-based privatization mechanism.

4. Moments Correction Due to the Perturbation

Based on the random variable X , the hypothesized distribution (initial or prior distribution around which the wavelet perturbation is introduced) has its k -th moment defined by

$$\mathbb{E}(X^k) := \int_{-\infty}^{+\infty} x^k dF_X(x), \quad (18)$$

providing its existence in the mathematical sense. By introducing the perturbation defined in (9), the new (adjusted/privatized) k -th moment is stated as

$$\mathbb{E}_{\text{priv}}(X^k) := \int_{-\infty}^{+\infty} x^k dF_{\text{priv}}(x). \quad (19)$$

Consider the equation given by $dF_{\text{priv}}(x) = dF_X(x) + \psi((b-a)F_X(x) + a)dF_X(x)$. Then, by using the expressions given in (18) and (19), it follows that

$$\mathbb{E}_{\text{priv}}(X^k) = \mathbb{E}(X^k) + \frac{1}{(b-a)} \int_a^b \left[F_X^{-1} \left(\frac{u-a}{b-a} \right) \right]^k \psi(u) du. \quad (20)$$

The second term on the right side of (20) accounts for a moment correction due to the introduced wavelet perturbation.

Let us consider now the particular case of a perturbation in a (normalized) uniform distribution, that is, $X \sim \mathcal{U}(0, 1)$. To evaluate the moments of the new CDF F_{priv} , under the wavelet perturbation ψ with a compact support $[0, 1]$, we have

$$\mathbb{E}_{\text{priv}}(X^k) := \mathbb{E}(X^k) + \int_0^1 u^k \psi(u) du. \tag{21}$$

Note that the moment of the wavelet used to build the additive perturbation also adds to the moment of the starting distribution, because

$$\mathbb{E}_{\text{priv}}(X^k) = \mathbb{E}(X^k) + \int_{-\infty}^{+\infty} u^k \psi(u) du = \mathbb{E}(X^k) + M_k. \tag{22}$$

If the support set is the unit interval, that is $[0, 1]$, then the formulas stated in (21) and (22) may be utilized. In the general case, if ψ has a support $[a, b] \neq [0, 1]$, we can build a modified (supported-normalized) wavelet defined as

$$\psi_{[0,1]} = \frac{\psi((b-a)x+a)}{(b-a)}.$$

Hence, we have that

$$\mathbb{E}_{\text{priv}}(X^k) = \mathbb{E}(X^k) + \int_{-\infty}^{+\infty} u^k \psi_{[0,1]}(u) du. \tag{23}$$

Under the assumption that the integral term given in (23) vanishes, the moments of the new and hypothesized distributions coincide.

5. Generalizing the Perturbation Approach at Further Levels

In the case that a beta perturbation occurs over a $\mathcal{U}[0, 1]$ distribution, it depends on its parameters α and θ of the perturbation wavelet. Thus, it is worth rewriting, via the equations stated in (1)–(9), that

$$F_{\text{priv}}(x) = \underbrace{x}_{\text{approximation}} + \underbrace{\Psi_{[0,1]}(x; \alpha, \theta)}_{\text{detail}}. \tag{24}$$

The interpretation presented in (24) of wavelet theory (approximation + detail) can be generalized into the lines of a wavelet tree with several levels. First, we present level-1 parameters (α, θ) by means of

$$F_{\text{level-1}}(x) = x + \Psi_{[0,1]}(x; \alpha, \theta). \tag{25}$$

In Figure 3, we can see examples of this case. Second, we introduce level-2 LH parameters $(\alpha_L, \theta_L; \alpha_H, \theta_H)$ considering

$$F_{\text{level-2}}(x) = \begin{cases} x + \Psi_{[0,1]}(2x; \alpha_L, \theta_L), & 0 \leq x \leq 1/2; \\ x + \Psi_{[0,1]}(2x - 1; \alpha_H, \theta_H), & 1/2 \leq x \leq 1. \end{cases} \tag{26}$$

An example can be provided using the parameters $\alpha_L = 4, \theta_L = 3$, and $\alpha_H = 3, \theta_H = 7$. These parameters are similar to those employed in Figure 3. However, note that different wavelets may be selected to fit different segments of the initial distribution support. For instance, in a level-2 perturbation, the sub-level-L can use a beta wavelet, whereas the sub-level-H may employ a Mexican-hat wavelet, denoted by $\Psi_{\hat{M}}$, as in Figure 5. The parameterization $\alpha_L = 4, \theta_L = 3$, and $\alpha_H = 3, \theta_H = 7$ is used in Figure 6, with the corresponding perturbation denoted by $\Psi_{\text{level-2}}$.

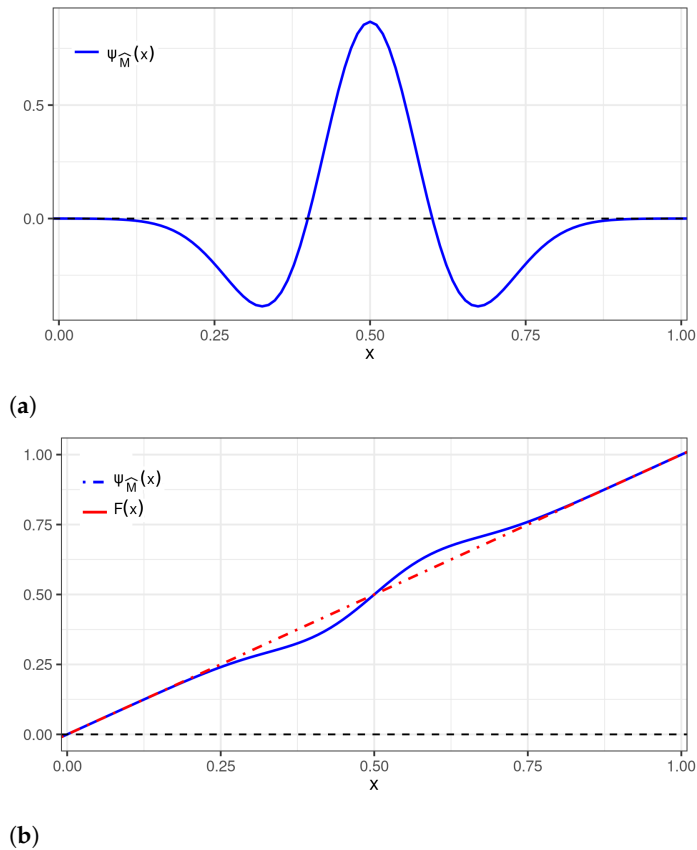


Figure 5. Plots of: (a) a Mexican-hat wavelet perturbation to be applied to the $\mathcal{U}[0,1]$ distribution; and (b) Mexican-hat wavelet perturbation (— blue) and uniform (- · - red) CDFs.

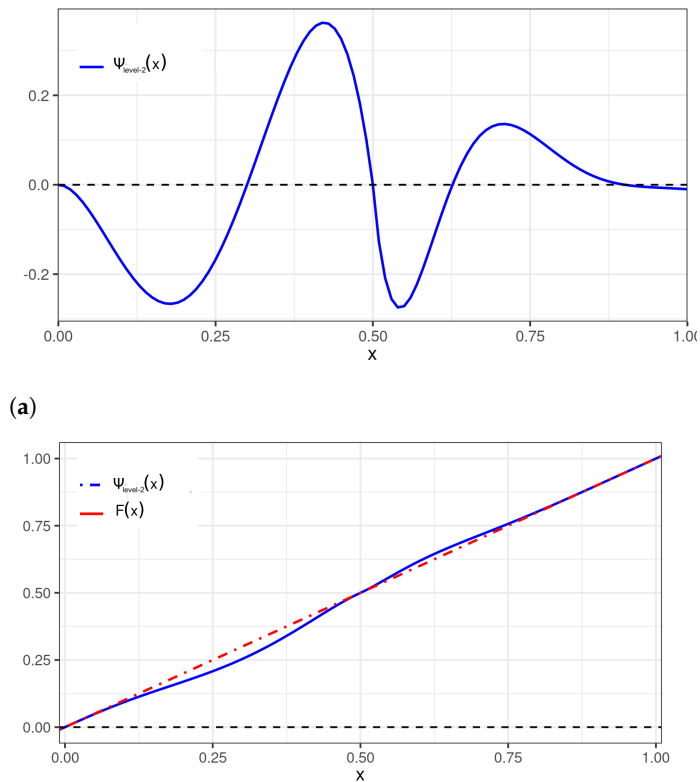


Figure 6. Plots of: (a) a level-2 beta wavelet perturbation to be applied to the $\mathcal{U}[0,1]$ distribution; and (b) level-2 beta wavelet perturbation (— blue) and uniform (- · - red) CDFs.

Next, we present level-4 LL LH HL HH parameters, $(\alpha_{LL}, \theta_{LL}; \alpha_{LH}, \theta_{LH}; \alpha_{HL}, \theta_{HL}; \alpha_{HH}, \theta_{HH})$ namely, stated as

$$F_{\text{level-4}}(x) = \begin{cases} x + \Psi_{[0,1]}(4x; \alpha_{LL}, \theta_{LL}), & 0 \leq x \leq 1/4; \\ x + \Psi_{[0,1]}(4x - 1; \alpha_{LH}, \theta_{LH}), & 1/4 \leq x \leq 1/2; \\ x + \Psi_{[0,1]}(4x - 2; \alpha_{HL}, \theta_{HL}), & 1/2 \leq x \leq 3/4; \\ x + \Psi_{[0,1]}(4x - 3; \alpha_{HH}, \theta_{HH}), & 3/4 \leq x \leq 1. \end{cases} \tag{27}$$

An example of this level-4 approach is illustrated utilizing the values given by

$$(\alpha_{LL}, \theta_{LL}; \alpha_{LH}, \theta_{LH}; \alpha_{HL}, \theta_{HL}; \alpha_{HH}, \theta_{HH}) = (4, 3; 3, 7; 5, 3; 2, 7).$$

An interpretation for this approach is considering a distinct perturbation in each quartile of the distribution such as:

- First quartile driven by $(\alpha_{LL}, \theta_{LL}) = (4, 3)$.
- Second quartile driven by $(\alpha_{LH}, \theta_{LH}) = (3, 7)$.
- Third quartile driven by $(\alpha_{HL}, \theta_{HL}) = (5, 3)$.
- Fourth quartile driven by $(\alpha_{HH}, \theta_{HH}) = (2, 7)$.

In short, the privatization mechanism allows us to perturb a probability distribution employing levels (applying a partition on the compact support), which may be very attractive when fitting data. We can use the expression stated in (25) when implementing one level, in (26) when implementing two levels, and in (27) when implementing four levels.

6. Empirical Application

Next, we apply our privatization approach to a real-world problem. An e-commerce company sells products on the Internet and wants to analyze the possibility of adding more servers or changing its most important server. By collecting daily data, we find many days in which the best server has almost all its hardware resources consumed 70% of the time. Looking at the empirical PDF and CDF, we see that a triangular distribution, with support on the set $[0, 1]$ and mode equal to 0.7, might represent the data well. However, when performing goodness-of-fit tests, the results tell us that a triangular distribution is not the best option. However, a “quasi-triangular” distribution could be an appropriate probability model for the random variable X that measures the daily proportion of times with full resource consumption of the best server. Among the known techniques to fit data, the privatization mechanism that we propose in this work is an excellent option to slightly perturb the triangular distribution and describe the data well. For the computational experiments, we utilize a database-sensor and two related algorithms.

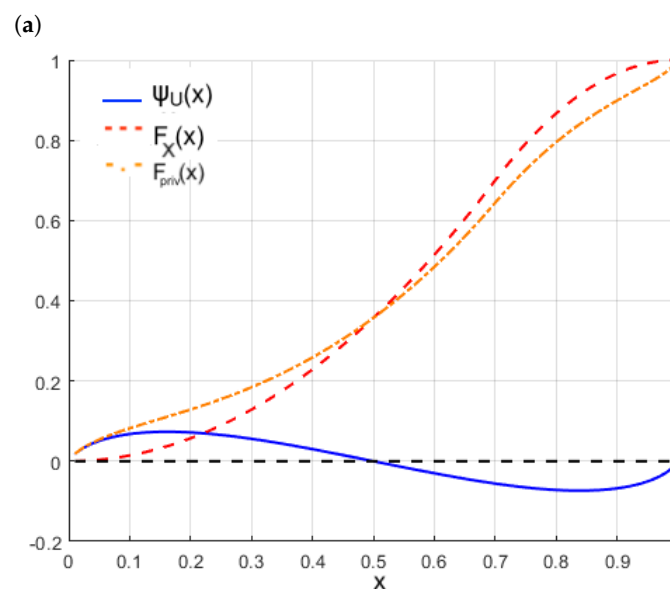
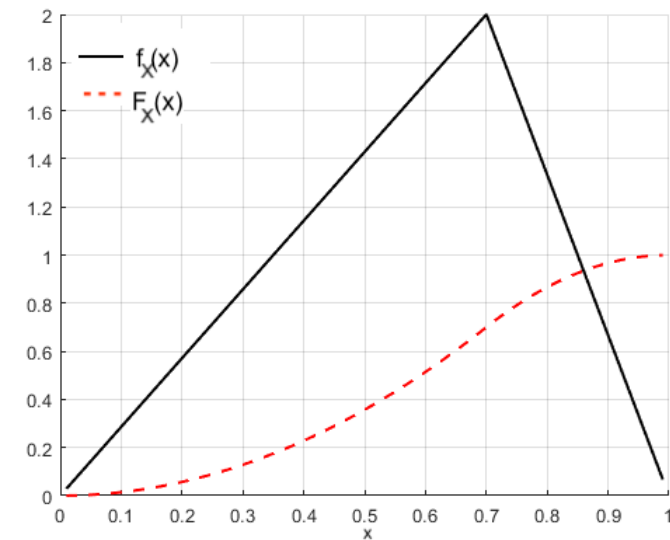
Let X be a continuous variable, which is triangularly distributed, with support on the interval $[0, 1]$, and whose mode is m , for $0 < m < 1$. The PDF and CDF of X are, respectively, given by

$$f_X(x) = \begin{cases} \frac{2x}{m}, & 0 \leq x \leq m; \\ \frac{2(1-x)}{1-m}, & m < x \leq 1; \end{cases}$$

and

$$F_X(x) = \begin{cases} \frac{x^2}{m}, & 0 \leq x \leq m; \\ 1 - \frac{(1-x)^2}{1-m}, & m < x \leq 1. \end{cases}$$

Now, we use the wavelet function defined in (13). Figure 7 shows the graphical plot of the CDF corresponding to X (original triangular distribution) and also the graphical plot of the privatized version that corresponds to the random variable X_{priv} (perturbed triangular distribution). We consider the value $m = 0.7$ in the calculations carried out. Note that, in the perturbed triangular distribution, the CDF values are greater than when compared to the original triangular distribution, for values of X less than 0.5, while for values of X greater than 0.5, the opposite occurs. This behavior is due to the wavelet function employed in such an empirical application. In practice, this method is flexible allowing us to choose the most convenient wavelet to fit the data.



(b)
Figure 7. Plots of: (a) PDF and CDF of the triangular distribution; and (b) wavelet perturbation (— blue), triangular (- - red), and perturbed triangular (- · - orange) CDFs.

For the computational experiments that were carried out, a database-sensor was used. Algorithm 1 shows the steps to perturb a probability distribution with compact support. If a perturbation by levels is required, we propose Algorithm 2 as a generalization of Section 5, where the number k of levels is left to the consideration of the data analyst.

Algorithm 1 Approach to perturb a probability distribution with a database-sensor.

- 1: Consider a random variable X with compact support $[a, b]$.
 - 2: Select a wavelet with compact support $[a, b]$ to perturb the distribution of the previous step, with the computations being performed by a first process denoted by A that sends the generated data to a database.
 - 3: State a sensor in the database that detects the entry of new data, so that, using a trigger, the sensor responds sending a copy of the stored data to a second process denoted by B.
 - 4: Establish that process B receives the perturbed data and is responsible for building the CDF of the resulting distribution.
 - 5: Confirm that process B generates the corresponding plots showing, between a and b , the original distribution, wavelet used, and perturbed distribution.
-

Algorithm 2 Approach to perturb a probability distribution by levels.

- 1: Select a probability distribution with compact support $[a, b]$.
 - 2: Apply a partition of k subintervals over the interval $[a, b]$ (not necessarily equispaced).
 - 3: Use Algorithm 1 on the interval $[a_i, b_i]$, for each i from 1 to k .
 - 4: Perform computations to unify the results on the interval $[a, b]$ of the previous step.
 - 5: Generate unified plots on the interval $[a, b]$ for the original distribution, wavelet used, and perturbed distribution.
-

7. Concluding Remarks

This paper has presented a new method for building an additive wavelet-based perturbation, as a privacy mechanism, to modify a given continuous probability distribution. Then, the initial guess could be perturbed as some sort of “prospecting within the ensemble of possible probability distributions around the starting distribution”.

The method we have proposed in this investigation is flexible with respect to the perturbation function that may be employed to fit the data, since different wavelets are available. A procedure was also offered to employ four different perturbations, one in each quartile of the original distribution, which can be quite attractive when fitting data. Examples of the proposed method were discussed. Computational experiments were carried out using a database-sensor and two related algorithms. Several knowledge areas can benefit from using the new method proposed in this study.

Stochastic programming, simulation studies, and multivariate analysis [31–34], among other areas of knowledge, may also benefit from the utilization of the new approach proposed in this investigation. The Internet of things, robotics, monitoring stations, telemetry, and the use of sensors are also important fields for data reading and fitting. Concrete applications via this new approach may now emerge, with an efficient configuration for the involved functions. Another benefit of this technique is its ease of implementation in any programming language. Software developers must be the first to get involved to make this technique available to data analysts. The areas of artificial intelligence, machine learning, and deep learning [35] constantly require new techniques for data fitting, whose areas are closely related to sensors. Accordingly, we think that the proposed privatization mechanism is an important contribution to increasing the spectrum of existing techniques. An avenue of future work to be considered is to provide a method that allows us to determine the most appropriate wavelet during data fitting.

Author Contributions: Conceptualization, H.M.d.O., R.O.; Data curation, H.M.d.O., R.O., C.M.-B.; Formal analysis, H.M.d.O., R.O., C.M.-B., V.L., C.C. Investigation, H.M.d.O., R.O., C.M.-B., V.L., C.C. Methodology, H.M.d.O., R.O., C.M.-B., V.L., C.C. Writing—original draft, H.M.d.O., R.O., C.M.-B., C.C. Writing—review and editing, V.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Council for Scientific and Technological Development (CNPq) through the grant number 305305/2019-0 (RO), and Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES), from the Brazilian government; and by FONDECYT, grant number 1200525 (V. Leiva), from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge, and Innovation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors warmly thank the editors and reviewers for their helpful comments which have led to an improved version of our paper.

Conflicts of Interest: There are no conflict of interest declared by the authors.

References

- Nor, A.K.M.; Pedapati, S.R.; Muhammad, M.; Leiva, V. Overview of explainable artificial intelligence for prognostic and health management of industrial assets based on preferred reporting items for systematic reviews and meta-analyses. *Sensors* **2021**, *21*, 8020. [[CrossRef](#)] [[PubMed](#)]
- Aykroyd, R.G.; Leiva, V.; Ruggeri, F. Recent developments of control charts, identification of big data sources and future trends of current research. *Technol. Forecast. Soc. Change* **2019**, *144*, 221–232. [[CrossRef](#)]
- Gomez-Deniz, E.; Leiva, V.; Calderin-Ojeda, E.; Chesneau, C. A novel claim size distribution based on a Birnbaum-Saunders and gamma mixture capturing extreme values in insurance: Estimation, regression, and applications. *Comput. Appl. Math.* **2022**, *41*, 171. [[CrossRef](#)]
- Bantan, R.A.R.; Jamal, F.; Chesneau, C.; Elgarhy, M. Truncated inverted Kumaraswamy generated family of distributions with applications. *Entropy* **2019**, *21*, 1089. [[CrossRef](#)]
- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; Wiley: New York, NY, USA, 1994; Volume 1.
- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; Wiley: New York, NY, USA, 1995; Volume 2.
- Tahir, M.H.; Cordeiro, G.M. Compounding of distributions: A survey and new generalized classes. *J. Stat. Distrib. Appl.* **2016**, *3*, 13. [[CrossRef](#)]
- Ahmad, Z.; Hamedani, G.G.; Butt, N.S. Recent developments in distribution theory: A brief survey and some new generalized classes of distributions. *Pak. J. Stat. Oper. Res.* **2019**, *10*, 87–110. [[CrossRef](#)]
- Aldahlan, M.A.; Jamal, F.; Chesneau, C.; Elgarhy, M.; Elbatal, I. The truncated Cauchy power family of distributions with inference and applications. *Entropy* **2020**, *22*, 346. [[CrossRef](#)] [[PubMed](#)]
- Bantan, R.A.R.; Chesneau, C.; Jamal, F.; Elbatal, I.; Elgarhy, M. The truncated Burr X-G family of distributions: Properties and applications to actuarial and financial data. *Entropy* **2021**, *23*, 1088. [[CrossRef](#)] [[PubMed](#)]
- Amigó, J.M.; Balogh, S.G.; Hernández, S. A brief review of generalized entropies. *Entropy* **2018**, *20*, 813. [[CrossRef](#)] [[PubMed](#)]
- Kotz, S.; Leiva, V.; Sanhueza, A. Two new mixture models related to the inverse Gaussian distribution. *Methodol. Comput. Appl. Probab.* **2010**, *12*, 199–212. [[CrossRef](#)]
- Alkadya, W.; ElBahnasy, K.; Leiva, V.; Gad, W. Classifying COVID-19 based on amino acids encoding with machine learning algorithms. *Chemom. Intell. Lab. Syst.* **2022**, *224*, 104535. [[CrossRef](#)] [[PubMed](#)]
- Nor, A.K.M.; Pedapati, S.R.; Muhammad, M.; Leiva, V. Abnormality detection and failure prediction using explainable Bayesian deep learning: Methodology and case study with industrial data. *Mathematics* **2022**, *10*, 554. [[CrossRef](#)]
- Balakrishnan, N.; Gupta, R.; Kundu, D.; Leiva, V.; Sanhueza, A. On some mixture models based on the Birnbaum-Saunders distribution and associated inference. *J. Stat. Plan. Inference* **2011**, *141*, 2175–2190. [[CrossRef](#)]
- Liu, Y.; Tang, S.; Liu, R.; Zhang, L.; Ma, Z. Secure and robust digital image watermarking scheme using logistic and RSA encryption. *Expert Syst. Appl.* **2018**, *97*, 95–105. [[CrossRef](#)]
- Kevorkian, J.; Cole, J.D. *Perturbation Methods in Applied Mathematics*; Springer: New York, NY, USA, 2013.
- Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press: Cambridge, UK, 1999.
- Burrus, C.S.; Gopinath, R.A.; Guo, H.; Odegaard, J.E.; Selesnick, I.W. *Introduction to Wavelets and Wavelet Transforms: A Primer*; Prentice Hall: New Jersey, NY, USA, 1998.
- Bae, C.; Lee, S.; Jung, Y. High-speed continuous wavelet transform processor for vital signal measurement using frequency-modulated continuous wave radar. *Sensors* **2022**, *22*, 3073. [[CrossRef](#)] [[PubMed](#)]
- Meyer, Y. *Ondelettes et Opérateur, I et II*; Hermann: Paris, France, 1990.
- Ghaderpour, E.; Pagiatakis, S.D.; Hassan, Q.K. A survey on change detection and time series analysis with applications. *Appl. Sci.* **2021**, *11*, 6141. [[CrossRef](#)]
- Qian, S. *Introduction to Time-Frequency and Wavelet Transforms*; Prentice-Hall Inc.: Upper Saddle River, NJ, USA, 2002.

24. Li, S.; Huang, W.; Shi, J.; Jiang, X.; Zhu, Z. A fast signal estimation method based on probability density functions for fault feature extraction of rolling bearings. *Appl. Sci.* **2019**, *9*, 3768. [[CrossRef](#)]
25. Dutra da Silva, R.; Robson, W.; Pedrini Schwartz, H. Image segmentation based on wavelet feature descriptor and dimensionality reduction applied to remote sensing. *Chilean J. Stat.* **2011**, *2*, 51–60
26. Daubechies, I. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **1988**, *41*, 906–996 [[CrossRef](#)]
27. Daubechies, I. *Ten Lectures on Wavelets*; CBMS-NSF Regional Conference Series in Applied Mathematics; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992; Volume 61.
28. de Oliveira, H.M.; Cintra, R.J. A new information theory concept: Information-weighted heavy-tailed distributions. *arXiv* **2016**, arXiv:1601.06412.
29. de Oliveira, H.M.; Araujo, G.A.A. Compactly supported one-cyclic wavelets derived from beta distributions. *J. Commun. Inf. Syst.* **2005**, *20*, 105–111. [[CrossRef](#)]
30. Vermehren, V.; de Oliveira, H.M. Close approximations for daubelets and their spectra. *arXiv* **2010**, arXiv:1502.01424.
31. Mahdi, E.; Leiva, V.; Mara'Beh, S.; Martin-Barreiro, C. A new approach to predicting cryptocurrency returns based on the gold prices with support vector machines during the COVID-19 pandemic using sensor-related data. *Sensors* **2021**, *21*, 6319. [[CrossRef](#)] [[PubMed](#)]
32. Rojas, F.; Leiva, V.; Huerta, M.; Martin-Barreiro, C. Lot-size models with uncertain demand considering its skewness/kurtosis and stochastic programming applied to hospital pharmacy with sensor-related COVID-19 data. *Sensors* **2021**, *21*, 5198. [[CrossRef](#)] [[PubMed](#)]
33. Ramirez-Figueroa, J.A.; Martin-Barreiro, C.; Nieto-Librero, A.B.; Leiva, V.; Galindo-Villardón, M.P. A new principal component analysis by particle swarm optimization with an environmental application for data science. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 1969–1984. [[CrossRef](#)]
34. Martin-Barreiro, C.; Ramirez-Figueroa, J.A.; Cabezas, X.; Leiva, V.; Martin-Casado, A.; Galindo-Villardón, M.P. A new algorithm for computing disjoint orthogonal components in the parallel factor analysis model with simulations and applications to real-world data. *Mathematics* **2021**, *9*, 2058. [[CrossRef](#)]
35. MacKay, D.J.C. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.