



Article Generalizing Normality: Different Estimation Methods for Skewed Information

Diego Carvalho do Nascimento ¹,*,[†][®], Pedro Luiz Ramos ^{2,†}[®], David Elal-Olivero ^{1,†}, Milton Cortes-Araya ^{1,†} and Francisco Louzada ^{2,†}[®]

- ¹ Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile; david.elal@uda.cl (D.E.-O.); milton.cortes@uda.cl (M.C.-A.)
- Institute of Mathematical Science and Computing, University of São Paulo, São Carlos 13566590, Brazil; pedrolramos@usp.br (P.L.R.); louzada@icmc.usp.br (F.L.)
- Correspondence: diego.nascimento@uda.cl
- + These authors contributed equally to this work.

Abstract: Normality is the most commonly used mathematical supposition in data modeling. Nonetheless, even based on the law of large numbers (LLN), normality is a strong presumption, given that the presence of asymmetry and multi-modality in real-world problems is expected. Thus, a flexible modification in the normal distribution proposed by Elal-Olivero adds a skewness parameter called Alpha-skew-normal (ASN) distribution, which enables bimodality and fat-tail, if needed, al-though it is sometimes not trivial to estimate this third parameter (regardless of the location and scale). This work analyzed seven different statistical inferential methods towards the ASN distribution on synthetic data and historical data of water flux from 21 rivers (channels) in the Atacama region. Moreover, the contributions of this paper are related to the estimations of probability surrounding rivers' flux levels in the surroundings of Copiapó city, which is the most economically important city of the third Chilean region and is known to be located in one of the driest areas on Earth (excluding the North and the South Poles). The results show the competitiveness of the MPS and RADE methods with respect to the MLE method, as well as their excellent performance.

Keywords: Alpha-skew-normal; bimodal distribution; asymmetry accommodation; water monitoring

1. Introduction

We live in the Big Data Era, where high volumes and many varieties of data characterization are often noticeable in data lakes. Despite the amount of observation, symmetry and smooth tails are not always observed. These characteristics are natural because we all live in a complex world with nonlinear relations and outliers that describe extreme values, which are more recurrent than easy statistical tools take into account. This new age requires flexible models and different reasoning based on data and information.

A question that is often asked in traditional departments worldwide is: "Are statistical methods becoming old fashioned?". Sir David Cox [1] explained that the focus is on the relevance and quality of data based on their coverage and representativeness, which provide confidence in the results in spite of the amount of information (large volumes) in a set, which may hold some potentially biased estimates with measurement errors. Efron and Hastie [2] discussed the relation between computer-related inferences and statistical inferences as a system of mathematical logic for guidance and correction that is complemented by large-scale prediction algorithms that are suitable for this new century.

Therefore, complexity is intrinsic to massive amounts of data, where high dimensionality and dynamism are often present [3,4]. Nonetheless, all of the information contained in the acquired data can be extracted by using an estimation method, e.g., with maximum likelihood estimation (MLE); a parametric version of such a method will be supported by a supposed distribution. Parametric approaches can be used to easily interpret patterns by



Citation: Nascimento, D.C.d.; Ramos P.L.; Elal-Olivero D.; Cortes-Araya M.; Louzada, F. Generalizing Normality: Different Estimation Methods for Skewed Information. *Symmetry* **2021**, *13*, 1067. https://doi.org/10.3390/ sym13061067

Academic Editor: Giuseppe Bagliesi

Received: 6 May 2021 Accepted: 10 June 2021 Published: 15 June 2021

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



Copyright: © 2021 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/). using parameters, enable association across variables, present a low computational cost, and are easier to implement in decision-making systems.

In many cases, the standard MLE may not return desirable results. Other estimation methods that return accurate estimates have been considered, such as estimators based on the least-square function [5], the product of spacing [6,7], or goodness-of-fit statistics [8]. There is not a unique method that performs better in all models, and the performance may depend on the selected parametric form [9,10]. Thus, it is necessary to use an efficient estimation method jointly with a flexible parametric model that covers many data patterns and that sometimes accommodates the asymmetry and multi-modality that may be contained in the data under consideration.

This is the case of meteorological data, which show significant changes, as well as a complex dynamic [11,12]. This field demands extra attention on data-driven models, which are needed in order to incorporate space–time dependence [13], structural changes [14], and extreme values [15]. Moreover, in a parametric world, a model supported by a probabilistic model (which deals with asymmetry and multi-modality [16,17]) is necessary. Thus, this paper was motivated by a case study of the water flux of 21 rivers (channels) in the surroundings of Copiapó city, which is located in the Atacama Desert and is one of the planet's driest areas. Moreover, we intend to exemplify evidence of the probability densities associated with these events' empirical distribution by using seven different approaches to statistical inference.

This paper is divided into four parts. Section 2 presents the motivation and details regarding the analyzed data. Section 3 provides the background of the adopted methodology with respect to the elements of statistical inference. Then, Sections 4 and 5 show the developments related to the methods when implemented on synthetic data and for the analysis of real-world data. Finally, Section 6 discusses the findings based on the obtained results.

2. The Data

The dataset adopted is related to fluviometric records (average monthly flows) from the Atacama Desert region (third-largest region of Chile) in the surroundings of Copiapó city. The historical period of these data is that of the ten years from January 2011 to December 2020; the data are associated with 21 rivers (or stream channels) and were obtained from the Chilean government website called *Direccion General de Aguas (Información Oficial Hidrometeorológica y de Calidad de Aguas en Línea*).

Historical events revealed the high periodicity of the low water flux of the region. However, cyclical events were also noticeable (such as two rainfall events and defrosting of glaciers/snow in summer, amongst others), creating the expected multi-modality and a large leptokurtosis.

During the processing of information, a decision-support system (DSS) sifts through and analyzes massive amounts of data, compiling comprehensive information that can be used in solving problems and making decisions. Figure 1 presents a flowchart of knowledge discovery in databases (KDD); the process involves information retrieval (IR), the decision-support system (DSS), and monitoring and forecasting.

Thus, the Atacama Desert watershed problem is part of a multi-dimensional study related to the analysis of a circular economy. It is essential to mention that uncertainty is always presented globally (as a measurement error or sample bias, amongst others). Nonetheless, probabilistic reasoning allows one to generalize results through statistical inference procedures.



Figure 1. Visual summary of the role of probabilistic reasoning in knowledge discovery in databases as a cornerstone for the quantification of uncertainty. Statistical inference procedures enable us to draw conclusions based on a sample and generalize them to an entire population.

3. Statistical Inference Elements

3.1. Alpha-Skew-Normal (ASN) Distribution

Let X be a random variable that follows an Alpha-skew-normal (ASN) distribution [18]; then, its probability density function (PDF) is given by

$$f(x|\alpha) = \frac{(1 - \alpha x)^2 + 1}{2 + \alpha^2} \phi(x),$$
(1)

where $x \in \mathbb{R}$, $\alpha \in \mathbb{R}$, and $\phi(\cdot)$ is the PDF of the standard normal distribution.

The cumulative density function (CDF) is given by

$$F(x|\alpha) = \Phi(x) + \alpha \left(\frac{2-\alpha x}{2+\alpha^2}\right)\phi(x).$$
(2)

Then, wrapping the ASN density $f(x|\alpha)$ with the parameters for location (μ) and scale (σ), that is, the random variable *T* is defined by $T = \mu + \sigma X$ for $\mu \in \mathbb{R}$ and $\sigma > 0$, this value is given by:

$$f(t|\mu,\sigma,\alpha) = \frac{(1-\alpha(t-\mu)\sigma^{-1})^2 + 1}{(2+\alpha^2)\sigma} \phi\left(\frac{t-\mu}{\sigma}\right).$$
(3)

Figure 2 presents different forms of the PDF of the ASN distribution. The CDF related to Equation (4) is

$$F(t|\mu,\sigma,\alpha) = \Phi\left(\frac{t-\mu}{\sigma}\right) + \alpha\left(\frac{2\sigma - \alpha(t-\mu)}{(2+\alpha^2)\sigma}\right)\phi\left(\frac{t-\mu}{\sigma}\right).$$
(4)

Given its flexibility, the ASN distribution has been used in data modeling and has been adopted in different fields, such as astronomy [19], the modeling of wind speed [20], and for benchmark data [21]. Figure 2 presents different forms of the PDF of the ASN distribution, assuming, for instance, $\mu = 0$ (location), $\sigma = 1$ (scale), and other values for α , in order to show the presence of asymmetry and bimodality (incorporating different heights between the modalities).



Figure 2. The PDF f(t) of the ASN distribution, where *t* is a random variable, assuming $\mu = 0$ (location), $\sigma = 1$ (scale), and different values for α (skewness).

3.2. Different Estimation Methods for the ASN Distribution

In this subsection, we will discuss seven different estimation methods (maximum likelihood estimation, ordinary and weighted least-square estimate, method of the maximum product of spacings, Cramer–von Mises minimum distance estimators, Anderson–Darling estimators, and right-tail Anderson–Darling estimators) for the parameters (μ , σ , and α) of the ASN distribution. Table 1 describes the methods used and the authors that proposed these inferential procedures. A comparison that used the cited estimators was presented for other models [22–24].

Table 1. A summary of seven inferential estimation methods.

Estimation Method	Abbreviation	Created by
Maximum Likelihood Estimation	MLE	Fisher [25]
Ordinary Least-Square Estimate	LSQ	Swain et al. [5]
Weighted Least-Square Estimate	WLQ	Swain et al. [5]
Maximum Product of Spacings	MPS	Cheng & Amin [6]
Cramer-von Mises Estimators	CME	Macdonald [26]
Anderson–Darling Estimator	ADE	Boos [27]
Right-Tail Anderson–Darling Estimator	RADE	Luceno [8]

Note that Carl Friedrich Gauss introduced the LSQ in 1822, and it is one of the oldest estimation procedures, although it was only in the paper of Swain et al. [5] that this approach was discussed for a class of non-normal models, and it then became a standard reference when applied in different probability distributions. The additional details related to the estimation of the ASN distribution's parameters are presented in the following subsections.

3.2.1. Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is widely used in data analysis; Fisher's derivation of information inequality was first used for the analysis of variance, and later for the estimation of functions derived from Euler's relation for homogeneous functions. Despite the fact that historical records of this technique have been widely exposed and defended by Ronald A. Fisher (who likely gained visibility because of his epic fights with Egon S. Pearson), its reasoning dates back to the mid-1700s [28].

Let t_1, t_2, \dots, t_n be a sample of a random sample of size *n* from $F(t|\mu, \sigma, \alpha)$. Considering $z_i = \frac{(t_i - \mu)}{\sigma}$, the maximum likelihood estimators $\hat{\mu}_{MLE}$, $\hat{\sigma}_{MLE}$, and $\hat{\alpha}_{MLE}$ can be obtained by maximizing

$$L(\mu,\sigma,\alpha) = \frac{1}{(2+\alpha^2)^n \sigma^n} \prod_{i=1}^n \left((1-\alpha z_i)^2 + 1 \right) \phi(z_i),$$
(5)

with respect to μ , σ , and α . The log-likelihood function of (5) is given by

$$l(\mu,\sigma,\alpha) = \sum_{i=1}^{n} \log\left((1-\alpha z_i)^2 + 1\right) - n\log(2+\alpha^2) - n\log(\sigma) + \sum_{i=1}^{n}\log\phi(z_i).$$
 (6)

From the expressions $\frac{\partial}{\partial \mu} l(\mu, \sigma, \alpha) = 0$, $\frac{\partial}{\partial \sigma} l(\mu, \sigma, \alpha) = 0$, $\frac{\partial}{\partial \alpha} l(\mu, \sigma, \alpha) = 0$, the likelihood equations are

$$\sum_{i=1}^{n} \frac{2\alpha(1-\alpha z_i)}{(1-\alpha z_i)^2+1} + \sum_{i=1}^{n} z_i = 0,$$
(7)

$$\sum_{i=1}^{n} \frac{2\alpha z_i (1 - \alpha z_i)}{(1 - \alpha z_i)^2 + 1} + \sum_{i=1}^{n} z_i^2 = 0,$$
(8)

$$\sum_{i=1}^{n} \frac{2z_i(1-\alpha z_i)}{(1-\alpha z_i)^2+1} + \frac{2n\alpha}{1+\alpha^2} = 0.$$
(9)

Numerical methods, such as the Newton–Rapshon, method are required in order to find the solution of a nonlinear system. Under mild conditions, the MLEs are asymptotically normally distributed with a joint multivariate normal distribution that is given by

$$(\hat{\mu}_{MLE}, \hat{\sigma}_{MLE}, \hat{\alpha}_{MLE}) \sim N_3 \Big[(\mu, \sigma, \alpha), I^{-1}(\mu, \sigma, \alpha)) \Big] \text{ as } n \to \infty,$$
 (10)

where $I(\mu, \sigma, \alpha)$ is the Fisher information matrix given by Elal-Olivero [18], who discussed that the ASN family satisfies these mild conditions.

This methodology is often adopted because of the sufficiency of the MLE, combined with its consistency (which leads to the statistical efficiency) and the fact that asymptotic normality is guaranteed. Next, we will present a series of minimum distance estimations (which are easily applied to estimate consistently unknown parameters) that are designed to reflect the proposed model in order to reproduce the probabilistic structure of the real-world phenomenon under study [29]. Minimum distance estimations provide consistent parameter estimates and are competitive, especially where other methods do not succeed.

3.2.2. Ordinary and Weighted Least-Square Estimates

Consider a random sample of order statistics of size *n* as $t = \{t_{(1)}, t_{(2)}, \dots, t_{(n)}\}$, where $t_{(i)}$ is not independent of $t_{(i+1)}$, and $F(t|\mu, \sigma, \alpha)$ is its monotonic function. The least-square (LSQ) estimators $\hat{\mu}_{LSE}$, $\hat{\sigma}_{LSE}$, and $\hat{\alpha}_{LSE}$ for the ASN distribution can be obtained by minimizing the parameters μ, σ , and α :

$$V(\mu,\sigma,\alpha) = \sum_{i=1}^{n} \left[F\left(t_{(i)} \mid \mu,\sigma,\alpha\right) - \frac{i}{n+1} \right]^2.$$
(11)

Thus, the LSQ equations can be obtained by solving the non-linear equations

$$\sum_{i=1}^{n} \left[F\left(t_{(i)} \mid \mu, \sigma, \alpha\right) - \frac{i}{n+1} \right] \Delta_{j}\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = 0, \quad j = 1, 2, 3,$$
(12)

where

$$\Delta_{1}\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = \frac{\partial}{\partial \mu} F\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = \frac{\phi(z_{i})}{(2 + \alpha^{2})\sigma} [2\alpha z_{i} - \alpha^{2} z_{i}^{2} - 2],$$

$$\Delta_{2}\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = \frac{\partial}{\partial \sigma} F\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = \frac{z_{i}\phi(z_{i})}{\sigma(2 + \alpha^{2})} [2\alpha z_{i} - \alpha^{2} z_{i}^{2} - 2],$$

$$\Delta_{3}\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = \frac{\partial}{\partial \alpha} F\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = \frac{\phi(z_{i})}{(2 + \alpha^{2})^{2}} [2 - 2\alpha^{2} - 4\alpha z_{i}].$$
(13)

Alternative solutions are obtained through high-precision numerical approximations of Δ_j for j = 1, 2, 3 partial derivatives.

Alternatively, the weighted least-squares (WLQ) estimates have been proposed for whenever an efficient method is required with small sets of data. $\hat{\mu}_{WLSE}$, $\hat{\sigma}_{WLSE}$, and $\hat{\alpha}_{WLSE}$ can be obtained by adopting the following minimized equation:

$$W(\mu,\sigma,\alpha) = \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[F\left(t_{(i)} \mid \mu,\sigma,\alpha\right) - \frac{i}{n+1} \right]^2.$$
(14)

The solutions deviate from those of the non-linear equations:

$$\sum_{i=1}^{n} \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left[F\left(t_{(i)} \mid \mu, \sigma, \alpha\right) - \frac{i}{n+1} \right] \Delta_j\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = 0, \quad j = 1, 2, 3,$$
(15)

where $\Delta_1(\cdot \mid \mu, \sigma, \alpha)$, $\Delta_2(\cdot \mid \mu, \sigma, \alpha)$ and $\Delta_3(\cdot \mid \mu, \sigma, \alpha)$ are given in (13). The WLQ estimation technique is particularly useful whenever one aims to weigh observations in proportion to the equivalence of the error variance for an observation in order to overcome the issue of non-constant variance.

3.2.3. Method of the Maximum Product of Spacings

The maximum product of spacings (MPS) method is a powerful alternative to MLE for estimating unknown parameters of continuous univariate distributions; it aims to maximize the geometric mean of the spacings in the data (differences between the values of the cumulative distribution function at neighborhood data points). Cheng and Amin proposed this method [6,30]—though it was also found independently by Ranneby [7]—as a Kullback–Leibler information approximation measurement. Some desirable properties of the MPS methods, such as their asymptotic efficiency, invariance, and, most importantly, consistency, are held more broadly (under general conditions) than for MLEs [30].

Let us represent the differences between the values of the cumulative distribution functions on their neighborhood data points with the function $D_i(\mu, \sigma, \alpha) = F(t_{(i)} | \mu, \sigma, \alpha) - F(t_{(i-1)} | \mu, \sigma, \alpha)$ for $i = \{1, 2, ..., n + 1, ...\}$ as a uniform spacing of a random sample from the ASN distribution, which is defined by $F(t_{(0)} | \mu, \sigma, \alpha) = 0$ and $F(t_{(n+1)} | \mu, \sigma, \alpha) = 1$. The constraint of $\sum_{i=1}^{n+1} D_i(\mu, \sigma, \alpha) = 1$ holds. Thus, the MPS estimates $\hat{\mu}_{MPS}$, $\hat{\sigma}_{MPS}$, and $\hat{\alpha}_{MPS}$ can be obtained by maximizing the geometric mean of the spacings:

$$G_{ASN}(\mu,\sigma,\alpha) = \left[\prod_{i=1}^{n+1} D_i(\mu,\sigma,\alpha)\right]^{\frac{1}{n+1}},$$
(16)

considering the maximization of this function (G_{ASN}) by adopting its logarithm as

$$H_{ASN}(\mu,\sigma,\alpha) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i(\mu,\sigma,\alpha).$$
(17)

The estimates of the unknown parameters $\hat{\mu}_{MPS}$, $\hat{\sigma}_{MPS}$, and $\hat{\alpha}_{MPS}$ are obtained by solving the nonlinear equations

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(\mu,\sigma,\alpha))}\Big[\Delta_j(t_{(i)}|\mu,\sigma,\alpha) - \Delta_j(t_{(i-1)}|\mu,\sigma,\alpha)\Big] = 0, \quad j = 1, 2, 3,$$
(18)

where $\Delta_1(\cdot \mid \mu, \sigma, \alpha)$, $\Delta_2(\cdot \mid \mu, \sigma, \alpha)$, and $\Delta_3(\cdot \mid \mu, \sigma, \alpha)$ are given, respectively, in Equation (13). It is important to mention that if $t_{(i+k)} = t_{(i+k-1)} = \ldots = t_{(i)}$, then $D_{i+k}(\mu, \sigma, \alpha) = 0$

 $D_{i+k-1}(\mu,\sigma,\alpha) = \ldots = D_i(\mu,\sigma,\alpha) = 0$. Therefore, the MPS estimators are sensitive to closely spaced observations—especially ties. When the ties are due to multiple observations, $D_i(\mu,\sigma,\alpha)$ should be replaced with the corresponding likelihood, $f(t_{(i)},\mu,\sigma,\alpha)$, because $t_{(i)} = t_{(i-1)}$.

Under mild conditions, for the ASN distribution, the MPS estimators are asymptotically normally distributed with a joint trivariate normal distribution given by

$$(\hat{\mu}_{MPS}, \hat{\sigma}_{MPS}, \hat{\alpha}_{MPS}) \sim N_3 \Big[(\mu, \sigma, \alpha), I^{-1}(\mu, \sigma, \alpha)) \Big] \text{ as } n \to \infty.$$
 (19)

3.2.4. The Cramer-von Mises Minimum Distance Estimators

Alternatively, an estimator that requires no assumptions about the distributions' parametric form, the Cramer–von Mises estimator (CME), is based on the difference between the estimates of the cumulative distribution function and the empirical distribution function [31,32]. These estimators operate based on the minimum distance across the "true" distribution (observed) and the "modeled" distribution (adjusted) through the maximum goodness of fit.

Macdonald [26] showed that the bias of the estimator in the CME presents smaller distances than those of other minimum distance estimators. The Cramer–von Mises estimates $\hat{\mu}_{CME}$, $\hat{\sigma}_{CME}$, and $\hat{\alpha}_{CME}$ of the parameters μ , σ , and α are obtained through minimization with

$$C(\mu,\sigma,\alpha) = \frac{1}{12n} + \sum_{i=1}^{n} \left(F\left(t_{(i)} \mid \mu,\sigma,\alpha\right) - \frac{2i-1}{2n} \right)^{2}.$$
 (20)

Thus, these estimates are also obtained by solving the non-linear equations:

$$\sum_{i=1}^{n} \left(F\left(t_{(i)} \mid \mu, \sigma, \alpha\right) - \frac{2i-1}{2n} \right) \Delta_{j}\left(t_{(i)} \mid \mu, \sigma, \alpha\right) = 0, \quad j = 1, 2, 3,$$
(21)

where $\Delta_1(\cdot \mid \mu, \sigma, \alpha)$, $\Delta_2(\cdot \mid \mu, \sigma, \alpha)$, and $\Delta_3(\cdot \mid \mu, \sigma, \alpha)$ are given, respectively, in Equation (13).

3.2.5. The Anderson–Darling and Right-Tail Anderson–Darling Estimators

Another type of minimum distance estimator is based on Anderson–Darling statistics, and it is often called an Anderson–Darling estimator (ADE). This estimator is based on the minimum distance estimation obtained by sampling data sorted in ascending order from the observed set (Y), and then X = Sort(Y), in combination with the permutation of $\{1, 2, ..., n\}$, which causes the X series to be sorted. Thus, this process is associated with the cumulative distribution function $F(\cdot)$ and the survival function $S(\cdot) = 1 - F(\cdot)$ for any PDF. In contrast, samples are only drawn from a uniform distribution if Y (and X) are samples from the PDF distribution.

The Anderson–Darling estimates $\hat{\mu}_{ADE}$, $\hat{\sigma}_{ADE}$, and $\hat{\alpha}_{ADE}$ of the parameters μ , σ , and α are obtained by minimizing, with respect to μ , σ , and α , the function:

$$A(\mu,\sigma,\alpha) = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \Big(\log F\Big(t_{(i)} \mid \mu,\sigma,\alpha\Big) + \log S\Big(t_{(n+1-i)} \mid \mu,\sigma,\alpha\Big) \Big).$$
(22)

These estimates can also be obtained by solving the non-linear equations:

$$\sum_{i=1}^{n} (2i-1) \left[\frac{\Delta_j \left(t_{(i)} \mid \mu, \sigma, \alpha \right)}{F \left(t_{(i)} \mid \mu, \sigma, \alpha \right)} - \frac{\Delta_j \left(t_{(n+1-i)} \mid \mu, \sigma, \alpha \right)}{S \left(t_{(n+1-i)} \mid \mu, \sigma, \alpha \right)} \right] = 0, \ j = 1, 2, 3.$$
(23)

Alternatively, one can improve the ADE's performance by considering the information held in the non-symmetrical differences between the theoretical CDF and the empirical CDF [33]. Thus, the right-tail Anderson–Darling estimator (RADE) is an alternative; $\hat{\mu}_{RADE}$, $\hat{\sigma}_{RADE}$, and $\hat{\alpha}_{RADE}$ of the parameters μ , σ , and α are obtained by minimizing the function:

$$R(\mu,\sigma,\alpha) = \frac{n}{2} - 2\sum_{i=1}^{n} F(t_{i:n} \mid \mu,\sigma,\alpha) - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \log S(t_{n+1-i:n} \mid \mu,\sigma,\alpha).$$
(24)

These estimates can also be obtained by solving the non-linear equations:

$$-2\sum_{i=1}^{n}\Delta_{j}(t_{i:n} \mid \mu, \sigma, \alpha) + \frac{1}{n}\sum_{i=1}^{n}(2i-1)\frac{\Delta_{j}(t_{n+1-i:n} \mid \mu, \sigma, \alpha)}{S(t_{n+1-i:n} \mid \mu, \sigma, \alpha)} = 0, \ j = 1, 2, 3,$$
(25)

where $\Delta_1(\cdot \mid \mu, \sigma, \alpha)$, $\Delta_2(\cdot \mid \mu, \sigma, \alpha)$, and $\Delta_3(\cdot \mid \mu, \sigma, \alpha)$ are given, respectively, in Equation (13).

4. Numerical Analysis

In this section, we investigated the behavior of the ASN distribution based on artificial (synthetic) data, as well as the modification of its parameters as a condition of the estimation method. Thus, a Monte Carlo simulation was carried out, the seven most commonly used estimation methods were considered for the parameters, and their efficiency was compared. The following approach was adopted. The procedure was:

- 1. Given a set of parameters from the $ASN(\mu, \sigma, \alpha)$ distribution, *N* samples of size *n* were generated;
- 2. For each generated set, based on the estimation methods (MLE, LSQ, WLQ, MPS, CME, ADE, and RADE), estimates of the parameters (μ , σ , and α) were calculated;
- 3. Then, considering $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\alpha})$ and $\theta = (\mu, \sigma, \alpha)$, the bias and mean squared error (MSE) of $\hat{\theta}$, which were given, respectively, by $\frac{1}{N} \sum_{k=1}^{N} \left(\hat{\theta}_{j}^{(k)} \theta_{j} \right)$ and $\frac{1}{N} \sum_{k=1}^{N} \left(\hat{\theta}_{j}^{(k)} \theta_{j} \right)^{2}$ for $j = \{1, 2, 3\}$ (each parameter), were computed. $\hat{\theta}_{j}^{(k)}$ denotes the estimate of θ_{j} obtained from sample k for $k = 1, 2, \cdots, N$.
- 4. The overall bias and the overall MSE were computed with $\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{3} (\hat{\theta}_{j}^{(k)} \theta_{j})$ and $\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{3} (\hat{\theta}_{j}^{(k)} \theta_{j})^{2}$.

The results of this simulation should return the estimation method that can be expected to be the most efficient based on if the estimations of both the bias and MSE are close to zero. For this simulation study, we adopted the R software [34], and for the maximization method, we used the *maxLik* and *stats4* packages [35]. The values chosen for the simulation parameters were N = 10,000 and $n = \{40, 60, 80, \dots, 300\}$. Due to the lack of space, we will present the results only for $\{\mu = 0, \sigma = 1, \alpha = 1\}$ and $\{\mu = 0, \sigma = 1, \alpha = 6\}$. Nonetheless, the following results are generalized by other choices of the vector of parameters θ . The estimation methods were considered under the same conditions in terms of samples, numbers of limit iterations, and initial values. Here, we considered the true values as initial values. However, we provide a simple approach that will be discussed in the next section in order to deal with real cases where good initial values are not available.

Figures 3 and 4 present the performance of the estimators in terms of the bias and MSE for the parameters μ , σ , and α when using the MLE, LSQ, WLQ, MPS, CME, ADE, and RADE with N = 10.000 simulated samples and different values of n. It can be observed that the MLE did not return adequate estimates for some parameter values and only

converged for samples of large sizes (moreover, an identifiability problem was noticeable when the MLE was adopted). These results show a drawback in the current approaches to obtaining parameter estimates with an ASN distribution. Although there is not a method that uniformly returns better estimates for all parameters and different parameter values, we observed that the MPS obtained the best results in terms of the minimum bias and MSE. Additionally, obtaining an estimate for α was quite challenging, but the MPS returned the smallest bias for this parameter. Therefore, we recommend using the MPS to obtain estimates for all parameters and parameter shows a distribution of the minimum bias and MSE.



Figure 3. Bias and MSE of the estimates of $\mu = 0$, $\sigma = 1$, and $\alpha = 1$ for N = 10,000 simulated samples of size *n* using the following methods: MLE, MPS, ADE, RADE, LSE, WLSE, and CME. Based on Figure 2, by choosing the configuration of these parameters ($\mu = 0.5$, $\sigma = 0.5$, and $\alpha = 3$), a bimodal PDF can be seen, which presents a larger peak to the left and a smaller peak to the right.



Figure 4. Bias and MSE of the estimates of $\mu = 0$, $\sigma = 1$, and $\alpha = 6$ for N = 10,000 simulated samples of size *n* using the following methods: MLE, MPS, ADE, RADE, LSE, WLSE, and CME.

5. Results

As we stated in Section 2, the motivation of this paper was driven by the water flux in the Atacama Desert or, more precisely, in the surroundings of Copiapó city. Figure 5 shows the empirical density of this phenomenon; a high concentration of low values is presented (near zero), although important events were also captured in this 10-year time window, such as a large amount of rainfall that caused a large leptokurtosis.



Figure 5. Empirical density function of the water flux in the 21 rivers/channels in the surroundings of Copiapó city. The solid gray shade represents the density (frequency) of each of the numerical records of the water flux, and the solid red line represents a smooth adjusted function.

Table 2 presents the statistical summaries (minimum, first quartile, median, mean, third quartile, and maximum) of the water flux by month. Because the weather is very constant in the region, the seasonality across years can be ignored, as low flux is common (close values through the minimum of the months). Nonetheless, the cycle in each month is essential; given events such as defrosting at the end of spring/beginning of summer (higher values in the third quartile in November and December), it is expected to that more water will be received in the system.

Month	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
JAN	0.02	0.06	0.31	0.5374	0.68	3.45	39
FEB	0.01	0.065	0.2	0.5165	0.6875	3.15	40
MAR	0.01	0.06	0.31	0.5449	0.85	3.24	37
APR	0.03	0.08	0.27	0.4494	0.5275	2.25	36
MAY	0.03	0.12	0.29	0.7859	0.55	19.47	47
JUN	0.02	0.12	0.35	0.9106	0.62	19.01	51
JUL	0.01	0.14	0.46	0.5636	0.64	2.58	53
AUG	0.01	0.1125	0.33	0.4692	0.6175	2.23	50
SEP	0.01	0.2025	0.45	0.5356	0.6775	2.66	52
OCT	0.02	0.1	0.37	0.5229	0.77	2.46	49
NOV	0.01	0.0775	0.365	0.5536	0.855	3.36	48
DEC	0.01	0.055	0.265	0.5639	0.7975	5.04	46

Table 2. Summary statistics of the water flux by month.

Logarithmic transformations have been used for a long time [36], though a normal distribution is often obtained. In some other situations, this is not the case; for instance, the dynamic of the water flux in the observed period, which is shown in Figure 6, shows

shown in Figure 7. It is essential to mention that the maximum historical values were in May and June (of 2017); they were related to the heavy rains that occurred in the region, which are also notable in the previous Table 2.



Figure 6. The empirical distribution of the log of the water flux and its frequency, which is represented by gray blocks. The black dashed line represents the adjusted ASN distribution based on the MPS ($\mu = -1.93, \sigma = 0.896, \alpha = -7.87$), which is represented by the blue dashed line, and the RADE ($\mu = -1.87, \sigma = 1.05, \alpha = -8.88$), which is represented by the red dot-dashed line.



Figure 7. The logarithm of the water flux dispersion records (y-coordinates) for each year (per panel) by month (x-coordinates).

The empirical distribution of this phenomenon is visualized in Figure 6, in which the dashed lines represents the adjusted density functions: the MPS in red and the RADE in blue. The initial values used to start the iteration procedures were obtained from

$$\tilde{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} \quad \text{and} \quad \tilde{\sigma} = \sum_{i=1}^{n} \frac{(x_i - \tilde{\mu})^2}{n},$$
(26)

while $\tilde{\alpha}$ was obtained from a grid search in the range (-10, -9.5, ..., 9, 10). To confirm that the MPS returned better estimates than the others, we used Ecdf-based goodness-of-fit tests of the ASN distribution model. Figure 8 shows a particular type of distance between the functions F_n and the observed values.



Figure 8. Ecdf-based test of the ASN distribution; RADE returned an AIC estimation of 274 and MPS returned an AIC estimation of 251.

After transforming the parameters with the exponential function, the practical implications of the obtained results are that, on average, it is expected that a monthly water flux of 0.1452 (from the MPS)—compared to 0.1541 (from the RADE)—will be observed, which will show a monthly dispersion of 2.45 (MPS) versus 2.85 (RADE) by considering the value of μ . Regarding the asymmetry estimated with both methods (α), they showed that it is expected that a frequent occurrence will be seen on the right side of the empirical distribution, that is, the recurrence of values greater than the average (μ) is expected. Based on the results found by Elal-Olivero [18], if an ASN distribution presents a parameter $\alpha > 1.34$, a bimodality will be obtained; if not, then shall only see a skewness. Through the transformation of the estimations towards the skewed parameter (α), it bimodality was obtained (in cases of both MPS and RADE); on the other hand, through the exponentiation of $\hat{\alpha}$, elements in the direction of skewed information were obtained. Figure 5 shows the presence of skewed data.

After confirming the ASN distribution's goodness of fit, the occurrence of an event could be associated with its density (or cumulative) probability function. For instance, extreme values can be seen; Table 3 shows some exemplifications that consider 1%, 10%, 50%, 99%, and 99.99%.

CUM Prob.	1%	10%	50%	99%	99.99%
Flux	0.0059	0.0174	0.3396	1.5068	16.281

Table 3. Cumulative event probability based on the adjusted ASN distribution (using RADE).

6. Conclusions

Uncertainty reveals a wide variety of processes and experiences that may follow different rules. However, other attributions of uncertainty, such as external (disposition) versus internal factors (ignorance), are assessed through statistical inference with certain philosophical interpretations of probability [37]. The utilities of each possible outcome lead to the choice of rational actions regardless of the observed results' uncertainty.

The example presented by this paper is the modeling of water flux, which is an essential element that has been placed under stress by significant population growth and an increase in the demand for water supply (for agriculture, industrial processes, mineral extraction, and human consumption) [38]. Therefore, planning the logistics of excessive water decisions according to their probabilistic distributions helps in unraveling this complex task [39,40]; this can be initiated by the analysis of the water flux, especially when environmental factors present limited sources for the water levels.

In the Atacama Desert, a north–south geographic band that is mainly located in northern Chile, precipitation is only a few millimeters per year, making it one of the driest places on Earth [41,42]. However, the vast expanses of the desert are punctuated by fertile valleys with rivers that originate in the central Andes and flow into the Pacific Ocean. Along these rivers, human populations have logically settled, exploiting this rare and precious water more and more throughout history, especially with the growing development of the mining industry, which is logically interested in the mineral resources of the Cordillera.

The hydrological regime of these rivers is characteristic of arid areas: Water flows from the peaks after the melting of the snowfall, glaciers, and permafrost located in the upper parts of the Cordillera [43]. In the context of climate change, it is therefore essential to understand the global hydrological cycles of these regions in order to set up a sustainable management policy [44]. This requires the implementation of tools for forecasting river flows, relative humidity, or any other water-related quantities, resulting in an inevitable need for in-depth knowledge of the physical phenomena that govern the entire hydrological cycle and, more precisely, the complex interactions among climate, ice, snow, and river flows.

Thus, this work proposed the investigation of different inference methods for the ASN probabilistic distribution, which is a promising and flexible distribution. Bimodality was noticeable and skewed information was observed in the historical series, but the distribution was nevertheless accommodated by the adopted probabilistic approach.

By using real-time analysis, big data solutions can now be implemented, as the process's probabilistic function was estimated and evidence of the goodness of fit was shown. In addition to the elements of the ASN distribution shown here, monitoring charts and other statistical process control (SPC) tools can also be explored because parametric distributions are often adopted. Future work should expand on the reasoning of the quantile estimations with explainable features (in a regression structure) that are associated with this problem, and forecasting may also be a motivation for future research.

Author Contributions: Conceptualization, methodology, software, writing—original draft preparation, D.C.d.N. and P.L.R.; validation, D.E.-O. and M.C.-A.; writing—review and editing, F.L., D.E.-O., and M.C.-A.; supervision and project administration, F.L. and D.E.-O.; funding acquisition, D.E.-O. and M.C.-A. All authors have read and agreed to the published version of the manuscript.

Funding: Diego Nascimento acknowledges the support from the São Paulo State Research Foundation (FAPESP process 2020/09174-5). Pedro L. Ramos acknowledges the support from the São Paulo State Research Foundation (FAPESP process 2017/25971-0). Francisco Louzada acknowledges the support from the São Paulo State Research Foundation (FAPESP Processes 2013/07375-0) and CNPq (grant no. 301976/2017-1).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available in Direccion General de Aguas at https://snia.mop.gob.cl/BNAConsultas/ (accessed on 4 May 2021). These data were derived from resources available in the public domain.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Cox, D.; Kartsonaki, C.; Keogh, R.H. Big data: Some statistical issues. Stat. Probab. Lett. 2018, 136, 111–115. [CrossRef]
- 2. Efron, B.; Hastie, T. Computer Age Statistical Inference; Cambridge University Press: Cambridge, UK, 2016; Volume 5.
- 3. Smith, J. Q. Decision Analysis: A Bayesian Approach; Chapman & Hall, Ltd.: London, UK, 1987.
- Leonelli, M.; Riccomagno, E.; Smith, J.Q. Coherent combination of probabilistic outputs for group decision making: An algebraic approach. OR Spectr. 2020, 42, 499–528. [CrossRef]
- 5. Swain, J.J.; Venkatraman, S.; Wilson, J.R. Least-squares estimation of distribution functions in johnson's translation system. *J. Stat. Comput. Simul.* **1988**, *29*, 271–297. [CrossRef]
- Cheng, R.; Amin, N. Maximum product of spacings estimation with application to the lognormal distribution. *Math. Rep.* 1979, 79, 1.
- 7. Ranneby, B. The maximum spacing method. an estimation method related to the maximum likelihood method. *Scand. J. Stat.* **1984**, *11*, 93–112.
- 8. Luceño, A. Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators. *Comput. Stat. Data Anal.* **2006**, *51*, 904–917. [CrossRef]
- 9. Louzada, F.; Ramos, P.L.; Ferreira, P.H. Exponential-poisson distribution: Estimation and applications to rainfall and aircraft data with zero occurrence. *Commun. Stat. Simul. Comput.* **2020**, *49*, 1024–1043. [CrossRef]
- 10. Ramos, P.L.; Nascimento, D.C.; Ferreira, P.H.; Weber, K.T.; Santos, T.E.; Louzada, F. Modeling traumatic brain injury lifetime data: Improved estimators for the generalized gamma distribution under small samples. *PLoS ONE* **2019**, *14*, e0221332. [CrossRef]
- 11. Bonnail, E.; Lima, R.C.; Turrieta, G.M. Trapping fresh sea breeze in desert? Health status of camanchaca, atacama's fog. *Environ. Sci. Pollut. Res.* **2018**, 25, 18204–18212. [CrossRef]
- Du, H.; Alexander, L.V.; Donat, M.G.; Lippmann, T.; Srivastava, A.; Salinger, J.; Kruger, A.; Choi, G.; He, H.S.; Fujibe, F.; et al. Precipitation from persistent extremes is increasing in most regions and globally. *Geophys. Res. Lett.* 2019, 46, 6041–6049. [CrossRef]
- 13. Lopes, H.F.; Salazar, E.; Gamerman, D. Spatial dynamic factor analysis. Bayesian Anal. 2008, 3, 759–792. [CrossRef]
- 14. Mutti, P.R.; Lúcio, P.S.; Dubreuil, V.; Bezerra, B.G. Ndvi time series stochastic models for the forecast of vegetation dynamics over desertification hotspots. *Int. J. Remote Sens.* 2020, *41*, 2759–2788. [CrossRef]
- 15. Dutfoy, A.; Parey, S.; Roche, N. Multivariate extreme value theory-a tutorial with applications to hydrology and meteorology. *Depend. Model.* **2014**, 2. [CrossRef]
- 16. Ramos, P.; Louzada, F. The generalized weighted lindley distribution: Properties, estimation and applications. *Cogent Math.* **2016**, 3, 1256022. [CrossRef]
- 17. Rodrigues, G.C.; Louzada, F.; Ramos, P.L. Poisson—Exponential distribution: Different methods of estimation. J. Appl. Stat. 2018, 45, 128–144. [CrossRef]
- 18. Elal-Olivero, D. Alpha-skew-normal distribution. Proyecciones 2010, 29, 224–240. [CrossRef]
- 19. Tarnopolski, M. Analysis of gamma-ray burst duration distribution using mixtures of skewed distributions. *Mon. Not. R. Astron. Soc.* **2016**, 458, 2024–2031. [CrossRef]
- Yang, K.; Aziz, M. Modeling Wind Speed Distributions Using Skewed Probability Functions: A Monte Carlo Simulation with Applications to Real Wind Speed Data. Available online: https://minds.wisconsin.edu/handle/1793/79304 (accessed on 4 May 2021).
- 21. Ara, A.; Louzada, F. The multivariate alpha skew gaussian distribution. *Bull. Braz. Math. Soc. New Ser.* 2019, 50, 823–843. [CrossRef]
- 22. Dey, S.; Kumar, D.; Ramos, P.L.; Louzada, F. Exponentiated chen distribution: Properties and estimation. *Commun. Stat. Simul. Comput.* **2017**, *46*, 8118–8139. [CrossRef]
- 23. Ramos, P.L.; Louzada, F.; Shimizu, T.K.; Luiz, A.O. The inverse weighted lindley distribution: Properties, estimation and an application on a failure time data. *Commun. Stat. Theory Methods* **2018**, *99*, 1–20. [CrossRef]
- 24. Teimouri, M.; Hoseini, S.M.; Nadarajah, S. Comparison of estimation methods for the Weibull distribution. *Statistics* **2013**, *47*, 93–109. [CrossRef]

- 25. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1922**, 222, 309–368.
- 26. Macdonald, P. An estimation procedure for mixtures of distribution. J. R. Stat. Soc. Ser. B 1971, 33, 326–329.
- 27. Boos, D.D. Minimum anderson-darling estimation. Commun. Stat. Theory Methods 1982, 11, 2747–2774. [CrossRef]
- 28. Stigler, S.M. The epic story of maximum likelihood. Stat. Sci. 2007, 22, 598–620. [CrossRef]
- 29. Wolfowitz, J. The minimum distance method. Ann. Math. Stat. 1957, 28, 75–88. [CrossRef]
- 30. Cheng, R.; Amin, N. Estimating parameters in continuous univariate distributions with a shifted origin. *J. R. Stat. Soc. Ser. B* **1983**, 45, 394–403. [CrossRef]
- 31. Cramér, H. On the composition of elementary errors: First paper: Mathematical deductions. *Scand. Actuar. J.* **1928**, 1928, 13–74. [CrossRef]
- 32. Von Mises, R. Statistik und Wahrheit; Julius Springer: Berlin/Heidelberg, Germany, 1928; Volume 20.
- 33. Ye, Y.; Lu, G.; Li, Y.; Jin, M. Unilateral right-tail anderson-darling test based spectrum sensing for cognitive radio. *Electron. Lett.* **2017**, *53*, 1256–1258. [CrossRef]
- 34. R Core Team. *R: A Language and Environment for Statistical Computing. (Version 3.3.1);* R Foundation for Statistical Computing: Vienna, Austria, 2014.
- 35. Henningsen, A.; Toomet, O. Maxlik: A package for maximum likelihood estimation in r. *Comput. Stat.* **2011**, *26*, 443–458. [CrossRef]
- 36. Finney, D. On the distribution of a variate whose logarithm is normally distributed. *Suppl. J. R. Stat. Soc.* **1941**, *7*, 155–161. [CrossRef]
- 37. Kahneman, D.; Tversky, A. Variants of uncertainty. Cognition 1982, 11, 143–157. [CrossRef]
- 38. Södergren, K.; Palm, J. How organization models impact the governing of industrial symbiosis in public wastewater management. an explorative study in sweden. *Water* **2021**, *13*, 824. [CrossRef]
- 39. Jain, A.; Ormsbee, L.E. Short-term water demand forecast modeling techniques—Conventional methods versus ai. *J. Am. Water Work. Assoc.* 2002, *94*, 64–72. [CrossRef]
- 40. Tu, Z.; Gao, X.; Xu, J.; Sun, W.; Sun, Y.; Su, D. A novel method for regional short-term forecasting of water level. *Water* **2021**, 13, 820. [CrossRef]
- 41. Bull, A.T.; Andrews, B.A.; Dorador, C.; Goodfellow, M. Introducing the Atacama Desert; Springer: Berlin/Heidelberg, Germany, 2018.
- 42. Grosjean, M.; Veit, H. Water Resources in the Arid Mountains of the Atacama Desert (Northern Chile): Past Climate Changes and Modern Conflicts; Springer: Dordrecht, The Netherlands, 2005.
- 43. Donoso, G.; Lictevout, E.; Rinaudo, J.-D. Groundwater management lessons from Chile. In *Sustainable Groundwater Management*; Springer: Berlin/Heidelberg, Germany, 2020.
- 44. Suárez, F.; Muñoz, J.; Fernández, B.; Dorsaz, J.-M.; Hunter, C.K.; Karavitis, C.A.; Gironás, J. Integrated water resource management and energy requirements for water supply in the Copiapó river basin, Chile. *Water* 2020, *6*, 2590–2613. [CrossRef]