Article | Received 26 June 2024; Accepted 30 September 2024; Published 14 October 2024
https://doi.org/10.55092/asymmetry20240004

The asymmetric cosine distribution

Christophe Chesneau

Department of Mathematics, LMNO, University of Caen-Normandie, 14032 Caen, France

* Correspondence author; E-mail: christophe.chesneau@unicaen.fr.

Abstract: In this article, we propose and study an improved version of the cosine distribution. Its main features are to be with support $[-1,1]$, to be mainly asymmetric by construction, and to be simple and flexible from a functional point of view, in particular, it retains the cosine feature that allows it to generate oscillatory shapes and multimodes, to unify the cosine distribution and the truncated exponential distribution, and to be used as a statistical model in different scenarios involving data contained in $[-1,1]$. We provide the main theory, including an expression of the key functions, quantiles and moments. Numerical work is carried out to determine the main measures of skewness and kurtosis. The practicality of the proposed distribution is illustrated with some data fitting scenarios, supported by graphical analysis.

Keywords: cosine distribution; asymmetric distribution; graphical analysis; reliability; parameter estimation

1. Introduction

Distributions with bounded support are useful in several areas of research and practical applications. They provide a balance between mathematical tractability and realistic modelling, especially when the underlying phenomena have natural limits. Well-known bounded support distributions include the beta distribution, the arcsine distribution, the Program Evaluation and Review Technique (PERT) distribution, the uniform distribution, the Irwin-Hall distribution, the Bates distribution, the logit-normal distribution, the Kumaraswamy distribution, the logit-metalog distribution, the Marchenko-Pastur distribution, the (raised) cosine distribution, the reciprocal distribution, the triangular distribution, the trapezoidal distribution, the truncated exponential distribution, the truncated normal distribution, the U-quadratic distribution, the von Mises distribution, and the continuous Bernoulli distribution. More details on the most famous of them can be found in [1].

Note that only a few of the distributions mentioned above belong to the family of symmetric distributions with support $[-1,1]$. The cosine distribution is an example, and perhaps one of the least studied, despite its high quality in several aspects. In order to convey the motivations of this article, a review of this singular distribution is necessary. As a first property, it is defined

with the following Probability Density Function (PDF):

$$f(x) = \begin{cases} \frac{1}{2}[1+\cos(\pi x)] & \text{if } x \in [-1,1], \\ \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

This PDF is symmetric about 0, *i.e.*, $f(x) = f(-x)$ for any $x \in \mathbb{R}$, it is bell-shaped and has support $[-1,1]$. For this reason, it is often considered a reasonable approximation of the PDF of the standard normal distribution, but over $[-1,1]$. The cosine distribution has applications in various fields such as signal processing and telecommunications. For the basics of it, we can refer to [2–6]. Recent developments include [7], which provides its main mathematical characterization, [8], which uses it to skew the normal distribution using the Azzalini scheme, and [9], which explores a generalization by adding two tuning parameters.

Compared to the other finite support distributions, this cosine distribution has not received much attention. In particular, we find no trace of an asymmetric version of it. Such a version is of interest because most data found on bounded intervals such as $[-1,1]$ do not satisfy the symmetry assumption, making the cosine distribution (and other symmetric distributions) inappropriate. The development of a simple asymmetric type of cosine distribution could therefore provide a more flexible and accurate modelling tool for data that exhibit skewness or other forms of asymmetry within $[-1,1]$, thus filling a gap in current statistical methods.

This article aims to fill this gap by proposing the asymmetric cosine (AC) distribution. It is defined by adapting an idea of [9] which consists in modulating the cosine term in the PDF of Equation (1) with a tuning parameter, and considering a multiplicative one-parameter exponential weight on the resulting function. This weight is the key to breaking the symmetry while maintaining a certain simplicity. Some advantages of the AC distribution are: (i) it is mainly asymmetric by construction, which is not so common for a distribution with support $[-1,1]$, (ii) it is simple from a functional point of view, while retaining the cosine feature that allows to generate oscillatory shapes and multimodes, (iii) it unifies the cosine distribution and the truncated exponential distribution (to be presented later), (iv) various estimation methods can be applied to it quite efficiently, and (v) it can then be used as a statistical model in different scenarios involving data contained in $[-1,1]$. In this article, we will support all these aspects with theoretical and practical work. In particular, several data fitting scenarios are examined to demonstrate the applicability of the AC distribution. The considered scenarios are simulated, except for the last two, which use famous real data sets of air quality measurements in New York and seismic events near Fiji.

The rest of the article consists of the following sections: Section 2 deals with the presentation and study of the AC distribution. Some applications are given in Section 3. A conclusion is given in Section 4.

2. The AC distribution

The definition of the AC distribution and its main characteristics are discussed in this section.

2.1. Definition

We start by defining a valid PDF with support $[-1,1]$. We take our inspiration from Equation (1) to derive an asymmetric cosine PDF.

Proposition 2.1 *The following function defines a valid PDF:*

$$f(x) = \begin{cases} K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{\beta x} & \text{if } x \in [-1,1], \\ \\ 0 & \text{if } x \notin [-1,1], \end{cases} \tag{2}$$

where $\alpha \in [-1,1]$, $\beta \in \mathbb{R}$ *and*

$$K_{\alpha,\beta} = \begin{cases} \dfrac{\beta(\pi^2+\beta^2)}{2[\pi^2+(1-\alpha)\beta^2]\sinh(\beta)} & \text{if } \beta \neq 0, \\ \\ \dfrac{1}{2} & \text{if } \beta = 0. \end{cases} \tag{3}$$

We recall that $\sinh(x) = (e^x - e^{-x})/2$, *for* $x \in \mathbb{R}$.

Proof. Clearly, for any $x \in \mathbb{R}$, $f(x)$ is continuous, except perhaps for one or two values, which are $x = -1$ and $x = 1$. For any $x \notin [-1,1]$ we have $f(x) = 0$. For any $x \in [-1,1]$, since $\alpha \in [-1,1]$ and $\cos(\pi x) \in [-1,1]$, we have

$$1 + \alpha\cos(\pi x) \geq 1 - |\alpha| \geq 0.$$

Furthermore, it is obvious that $e^{\beta x} \geq 0$ for any $\beta \in \mathbb{R}$. As a result we have $f(x) \geq 0$ for any $x \in \mathbb{R}$. Let us now prove that $\int_{-\infty}^{+\infty} f(x)dx = 1$. We have

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-1}^{1} f(x)dx = K_{\alpha,\beta}\int_{-1}^{1}[1+\alpha\cos(\pi x)]e^{\beta x}dx.$$

To go on, for any $a \in \mathbb{R}$ and $b \in \mathbb{R}$ with $(a,b) \neq (0,0)$, we need the following primitive formula:

$$\int^{x} \cos(at)e^{bt}dt = \frac{e^{bx}}{a^2+b^2}[b\cos(ax) + a\sin(ax)] + C, \tag{4}$$

where C denotes a certain generic constant. See [10, 2.663, item 3].

For any $\beta \neq 0$, by applying Equation (4) with $a = \pi$ and $b = \beta$, we obtain

$$\begin{aligned}
&\int_{-1}^{1}[1+\alpha\cos(\pi x)]e^{\beta x}dx = \int_{-1}^{1}e^{\beta x}dx + \alpha\int_{-1}^{1}\cos(\pi x)e^{\beta x}dx \\
&= \frac{1}{\beta}(e^{\beta}-e^{-\beta}) + \alpha\left\{\frac{e^{\beta}}{\pi^2+\beta^2}[\beta\cos(\pi)+\pi\sin(\pi)] - \frac{e^{-\beta}}{\pi^2+\beta^2}[\beta\cos(\pi)-\pi\sin(\pi)]\right\} \\
&= \frac{1}{\beta}(e^{\beta}-e^{-\beta}) - \frac{\alpha\beta}{\pi^2+\beta^2}(e^{\beta}-e^{-\beta}) = 2\left(\frac{1}{\beta} - \frac{\alpha\beta}{\pi^2+\beta^2}\right)\sinh(\beta) \\
&= \frac{2[\pi^2+(1-\alpha)\beta^2]\sinh(\beta)}{\beta(\pi^2+\beta^2)} = \frac{1}{K_{\alpha,\beta}}.
\end{aligned} \tag{5}$$

For the case $\beta = 0$, we directly obtain

$$\int_{-1}^{1}[1+\alpha\cos(\pi x)]dx = \int_{-1}^{1}dx + \alpha\int_{-1}^{1}\cos(\pi x)dx = \frac{1}{2} + \alpha\times 0 = \frac{1}{2} = \frac{1}{K_{\alpha,0}}.$$

As a result, we have $\int_{-\infty}^{+\infty} f(x)dx = 1$. All the conditions are met; $f(x)$ is a valid PDF. □

We define the AC distribution with parameters $\alpha \in [-1,1]$ and $\beta \in \mathbb{R}$ by the distribution with the PDF given in Equation (2). Thus, a random variable X having this distribution satisfies, for any $A \subseteq \mathbb{R}$, $P(X \in A) = \int_A f(x)dx$, where P denotes the probability operator.

Let us now make some distributional remarks. By taking $\alpha = 1$ and $\beta = 0$, the AC distribution reduces to the cosine distribution. Furthermore, by taking $\alpha = 0$, the PDF of the AC distribution becomes

$$f(x) = \begin{cases} K_{0,\beta}e^{\beta x} & \text{if } x \in [-1,1], \\ 0 & \text{if } x \notin [-1,1], \end{cases}$$

with $\beta \in \mathbb{R}$ and

$$K_{0,\beta} = \begin{cases} \dfrac{\beta}{2\sinh(\beta)} & \text{if } \beta \neq 0, \\ \dfrac{1}{2} & \text{if } \beta = 0, \end{cases}$$

which is the PDF of the truncated exponential distribution over $[-1,1]$. See [11]. Note that, for the special case $\alpha = 0$ and $\beta = 0$, we get the uniform distribution over $[-1,1]$. Combined with the above statement, the AC distribution can thus be seen as a compromise between the cosine distribution and the truncated exponential distribution over $[-1,1]$.

Based on the PDF, the role of the parameter α is to modulate the cosine term, while the role of β is to activate the exponential function, which is designed to break the symmetry of the original cosine distribution. Indeed, for any $\beta \neq 0$ and any $x \in [-1,1]/\{0\}$, it is clear that

$$f(-x) = K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{-\beta x} \neq K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{\beta x} = f(x).$$

Thus, thanks to the combined effects of α and β, the PDF of the AC distributions aims to benefit from great flexibility, including varying symmetric and asymmetric shapes. This is immediately apparent from a basic function analysis. In particular, we have the special value $f(0) = K_{\alpha,\beta}(1+\alpha)$, which shows a strong dependence on α and β. Furthermore, the possible extremum(a) of $f(x)$ obtained by solving $f'(x) = 0$, which gives the following complex equation:

$$\alpha\left[\beta\cos(\pi x) - \pi\sin(\pi x)\right] = -\beta.$$

Since $\alpha \in [-1,1]$ and $\beta \in \mathbb{R}$, with the oscillatory nature of the cosine and sine functions, the explicit solution(s) of this equation has(have) no analytical expression. To support the arguments of flexibility and asymmetry, we proceed with a graphical analysis using the functionalities of the software R (see [12]).

Figure 1 shows a summary of the possible curves of the PDF of the AC distribution for selected values of the parameters α and β.

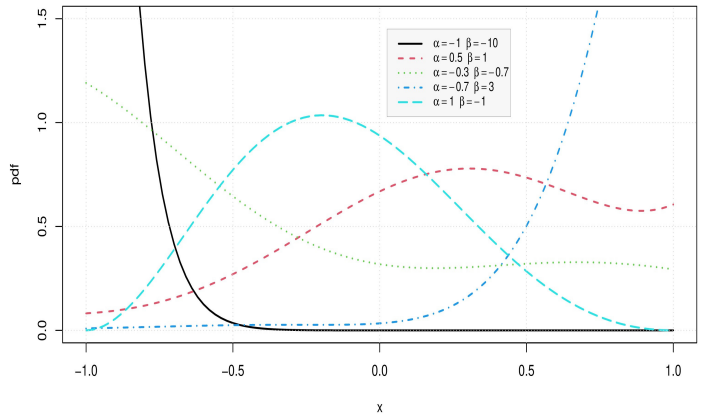

Figure 1. Some curves of the PDF of the AC distribution for selected values of the parameters α and β.

Many different shapes are observed. These include monotonic and non-monotonic shapes with some ripples.

To visualize the role of α and β in these variations, Figures 2, 3, 4 and 5 complete this by fixing one parameter, either α or β, and varying the other with positive values only on one hand and negative values on the other.

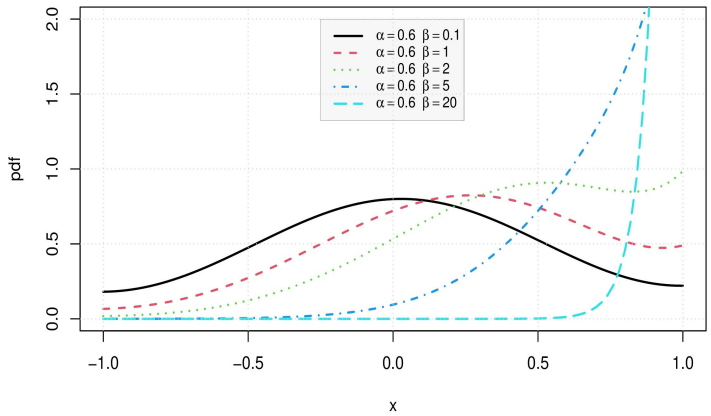

Figure 2. Some curves of the PDF of the AC distribution for $\alpha = 0.6$ and varying values of β.

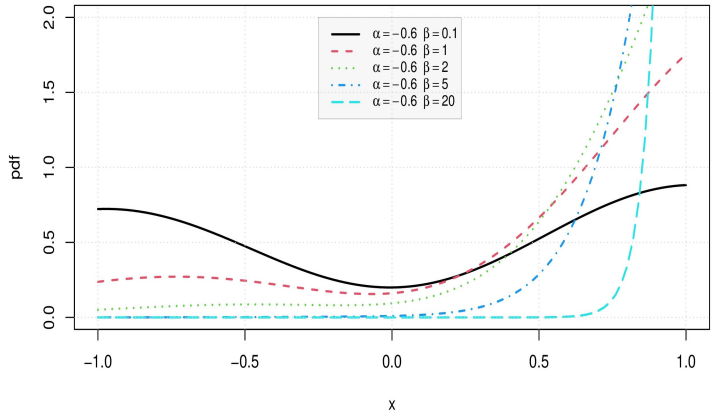

Figure 3. Some curves of the PDF of the AC distribution for $\alpha = -0.6$ and varying values of β.

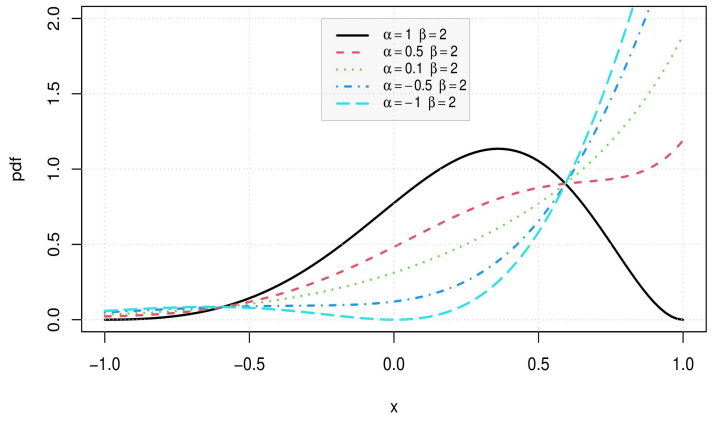

Figure 4. Some curves of the PDF of the AC distribution for $\beta = 2$ and varying values of α.

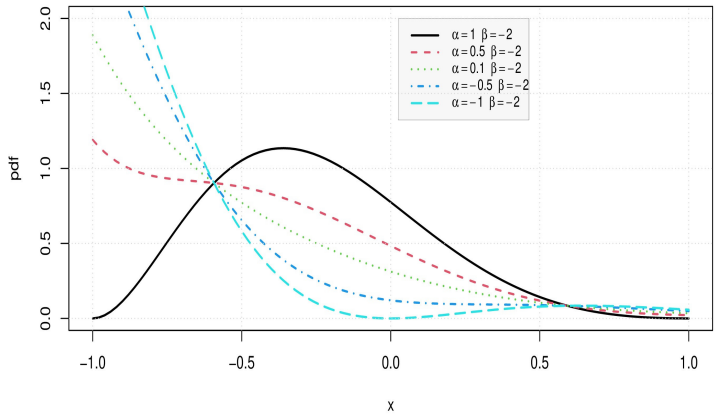

Figure 5. Some curves of the PDF of the AC distribution for $\beta = -2$ and varying values of α.

These figures show the flexibility of the AC distribution and also its rich panel of asymmetric shapes depending on the values of α and β. This demonstrates the success of our strategy of introducing varying asymmetry into the PDF of the cosine distribution.

We end this part by a secondary mathematical remark. The PDF of the AC distribution can be expressed in different summation forms, presented below. First, basically, for any

$x \in [-1,1]$, we can write

$$f(x) = K_{\alpha,\beta} e^{\beta x} + \alpha K_{\alpha,\beta} \cos(\pi x) e^{\beta x}.$$

We also can write

$$f(x) = f(x)|_{\alpha=1} + (\alpha - 1) K_{\alpha,\beta} \cos(\pi x) e^{\beta x}.$$

More technically, using the formula

$$[1 + \alpha \cos(\pi x)][1 - \alpha \cos(\pi x)] = 1 - \alpha^2 [\cos(\pi x)]^2 = 1 - \alpha^2 \{1 - [\sin(\pi x)]^2\},$$

we have

$$f(x) = K_{\alpha,\beta} \left\{ \frac{1 - \alpha^2}{1 - \alpha \cos(\pi x)} + \frac{\alpha^2 [\sin(\pi x)]^2}{1 - \alpha \cos(\pi x)} \right\}.$$

These diverse expressions can base the AC distributions on other existing distributions, or create new ones that extend the scope of the AC distribution.

2.2. Key functions

The Cumulative Distribution Function (CDF) associated with the AC distribution is determined in the result below.

Proposition 2.2 *The CDF of a random variable X having the AC distribution with parameters* $\alpha \in [-1,1]$ *and* $\beta \in \mathbb{R}$ *is given by*

$$F(x) = P(X \in (-\infty, x))$$

$$= \begin{cases} 0 & \text{if } x < -1 \\ K_{\alpha,\beta} \left[\frac{1}{\beta} (e^{\beta x} - e^{-\beta}) + \frac{\alpha}{\pi^2 + \beta^2} \left\{ e^{\beta x} [\beta \cos(\pi x) + \pi \sin(\pi x)] + \beta e^{-\beta} \right\} \right] & \text{if } x \in [-1,1], \\ 1 & \text{if } x > 1, \end{cases}$$

where $K_{\alpha,\beta}$ *is given as Equation* (3).

Proof. The value of $F(x)$ for $x < -1$ and $x > 1$ are immediate consequences of the support $[-1,1]$ of the AC distribution. For any $x \in [-1,1]$, we have

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} f(t)dt = K_{\alpha,\beta} \int_{-1}^{x} [1 + \alpha \cos(\pi t)] e^{\beta t} dt.$$

Let us now proceed as in the proof of Proposition 2.1, using Equation (4) with $a = \pi$ and

$b = \beta$. We have

$$\int_{-1}^{x} [1 + \alpha \cos(\pi t)] e^{\beta t} dt = \int_{-1}^{x} e^{\beta t} dt + \alpha \int_{-1}^{x} \cos(\pi t) e^{\beta t} dt$$
$$= \frac{1}{\beta}(e^{\beta x} - e^{-\beta}) + \alpha \left\{ \frac{e^{\beta x}}{\pi^2 + \beta^2} [\beta \cos(\pi x) + \pi \sin(\pi x)] - \frac{e^{-\beta}}{\pi^2 + \beta^2} [\beta \cos(\pi) - \pi \sin(\pi)] \right\}$$
$$= \frac{1}{\beta}(e^{\beta x} - e^{-\beta}) + \frac{\alpha}{\pi^2 + \beta^2} \left\{ e^{\beta x} [\beta \cos(\pi x) + \pi \sin(\pi x)] + \beta e^{-\beta} \right\}.$$

So, by multiplication with $K_{\alpha,\beta}$, we have

$$F(x) = K_{\alpha,\beta} \left[\frac{1}{\beta}(e^{\beta x} - e^{-\beta}) + \frac{\alpha}{\pi^2 + \beta^2} \left\{ e^{\beta x} [\beta \cos(\pi x) + \pi \sin(\pi x)] + \beta e^{-\beta} \right\} \right].$$

The desired expression is obtained, ending the proof. □

Figure 6 shows a summary of the possible curves of this CDF for the same selected values of the parameters α and β considered in Figure 1.

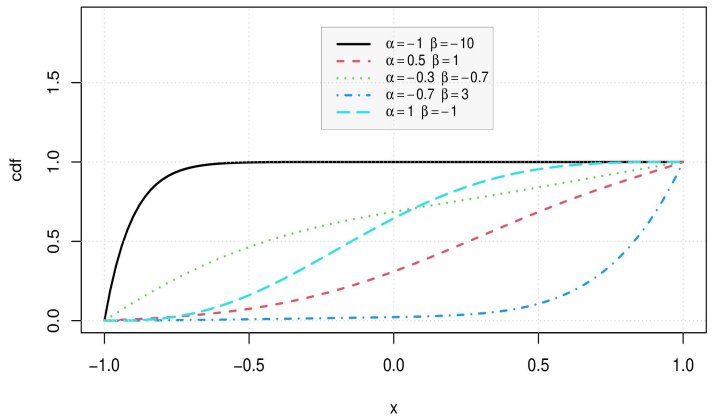

Figure 6. Some curves of the CDF of the AC distribution for selected values of the parameters α and β.

Clearly, various concave and convex shapes are observed, validating the flexible nature of the AC distribution.

Based on its CDF, we can examine the quantiles of the AC distribution. For any $p \in [0,1]$, the quantile of order p, denoted by x_p, satisfies the equation $F(x_p) = p$. Due to the complexity of $F(x)$, there is no closed-form expression for x_p. Therefore, a numerical study needs to be done. In the spirit of [9, Table 1], using the function uniroot of the software R, we provide a table of values for selected values of the parameters α and β (the same ones considered in Figure 1), and p. The results are given in Table 1.

Table 1. Some quantiles of order p of the AC distribution for selected values of the parameters α and β, and p.

$p \to$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\alpha=-1, \beta=-10$	−1.00	−0.99	−0.98	−0.97	−0.95	−0.93	−0.91	−0.89	−0.85	−0.79	1.00
$\alpha=0.5, \beta=1$	−1.00	−0.42	−0.18	−0.02	0.13	0.26	0.39	0.52	0.67	0.83	1.00
$\alpha=-0.3, \beta=-0.7$	−1.00	−0.91	−0.82	−0.71	−0.59	−0.44	−0.24	0.04	0.37	0.68	1.00
$\alpha=-0.7, \beta=3$	−1.00	0.49	0.63	0.72	0.78	0.83	0.87	0.91	0.94	0.97	1.00
$\alpha=1, \beta=-1$	−1.00	−0.59	−0.45	−0.34	−0.24	−0.15	−0.05	0.06	0.19	0.36	1.00

For example, considering $\alpha = 0.5$ and $\beta = 1$, and $p = 0.4$, we have $F(x_p) = p$ for $x_p = 0.13$. This table shows the large variation in the values of the quantiles, and also that they are quite calculable for all values of $\alpha \in [-1,1]$ and $\beta \in \mathbb{R}$. It is also shown that a deep numerical quantile analysis of the AC distribution is feasible, including the Bowley skewness and the Moors kurtosis. See [13] and [14], respectively.

Still based on the CDF, we can define some useful reliability functions. In particular, the survival function of the AC distribution is defined as

$$S(x) = 1 - F(x)$$
$$= \begin{cases} 1 & \text{if } x < -1 \\ 1 - K_{\alpha,\beta}\left[\dfrac{1}{\beta}(e^{\beta x} - e^{-\beta}) + \dfrac{\alpha}{\pi^2+\beta^2}\left\{e^{\beta x}[\beta\cos(\pi x) + \pi\sin(\pi x)] + \beta e^{-\beta}\right\}\right] & \text{if } x \in [-1,1], \\ 0 & \text{if } x > 1. \end{cases}$$

The Hazard Rate Function (HRF) of the AC distribution follows as

$$h(x) = \frac{f(x)}{S(x)}$$
$$= \begin{cases} 0 & \text{if } x < -1 \\ \dfrac{K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{\beta x}}{1 - K_{\alpha,\beta}\left[(1/\beta)(e^{\beta x} - e^{-\beta}) + [\alpha/(\pi^2+\beta^2)]\left\{e^{\beta x}[\beta\cos(\pi x) + \pi\sin(\pi x)] + \beta e^{-\beta}\right\}\right]} & \text{if } x \in [-1,1], \\ 1 & \text{if } x > 1. \end{cases}$$

We expect a certain variety of shapes for this function. This is confirmed in Figure 7, where some of these curves are shown for selected values of the parameters α and β (the same ones considered in Figure 1), with a small and a large view with respect to the y-axis to see all the details.

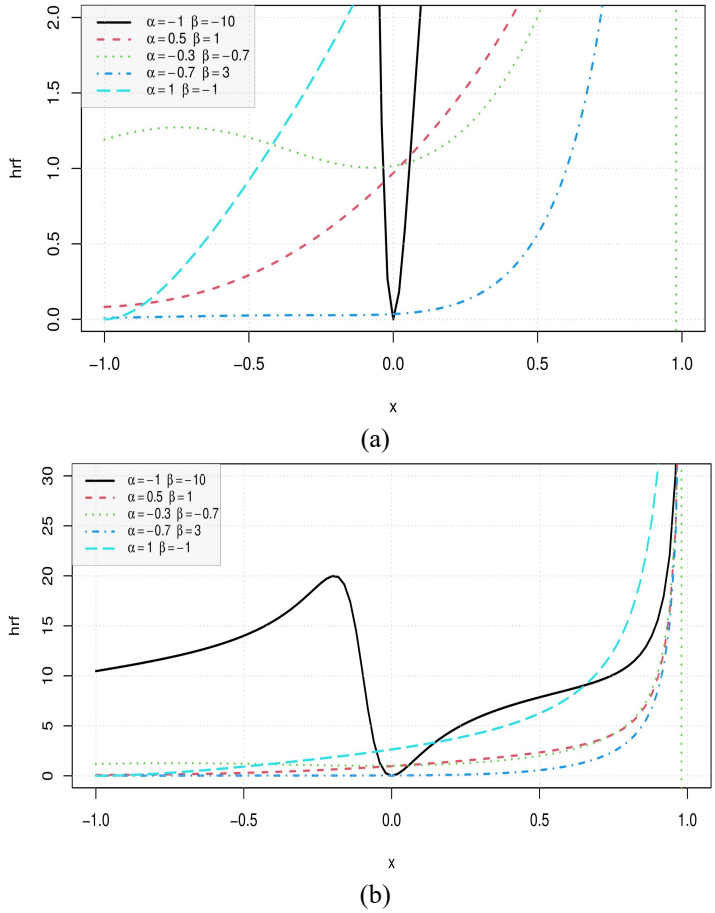

Figure 7. Some curves of the HRF of the AC distribution for selected values of the parameters α and β for (a) y-axis $\in [0,2]$ for zoom reasons, and (b) y-axis $\in [0,30]$.

The HRF corresponding to the singular black curve is typical of that obtained with trigonometric HRFs. See, for example, [15] in another distributional setting. The AC distribution can thus deal with standard and exotic HRFs inherent in data in $[-1,1]$.

2.3. Moments

Since the AC distribution has bounded support, most moment measures exist. With this in mind, we now examine the moment properties of the AC distribution, starting with the corresponding moment generating function determined in the proposition below.

Proposition 2.3 *The moment generating function of a random variable X having the AC distribution with parameters $\alpha \in [-1,1]$ and $\beta \in \mathbb{R}$ is given by*

$$G(s) = E(e^{sX}) = K_{\alpha,\beta} \frac{2[\pi^2 + (1-\alpha)(\beta+s)^2]\sinh(\beta+s)}{(\beta+s)[\pi^2+(\beta+s)^2]}, \quad s \in \mathbb{R}/\{-\beta\},$$

where E denotes the mathematical expectation operator and $K_{\alpha,\beta}$ is given as Equation (3).

Proof. We have

$$G(s) = \int_{-\infty}^{+\infty} e^{st} f(t)dt = K_{\alpha,\beta} \int_{-1}^{1} e^{st}[1+\alpha\cos(\pi t)]e^{\beta t}dt = K_{\alpha,\beta} \int_{-1}^{1} [1+\alpha\cos(\pi t)]e^{(\beta+s)t}dt.$$

This last integral term was almost calculated in Equation (5); it is enough to replace β by $\beta+s$.

We thus have

$$G(s) = K_{\alpha,\beta}\frac{1}{K_{\alpha,\beta+s}} = K_{\alpha,\beta}\frac{2[\pi^2+(1-\alpha)(\beta+s)^2]\sinh(\beta+s)}{(\beta+s)[\pi^2+(\beta+s)^2]}.$$

This ends the proof. □

The special case $s=-\beta$ will be considered in Proposition 2.6.

From this result, we easily derive the characteristic function of a random variable X having the AC distribution; it is given as

$$\varphi(t) = E(e^{itX}) = K_{\alpha,\beta}\frac{2[\pi^2+(1-\alpha)(\beta+it)^2]\sinh(\beta+it)}{(\beta+s)[\pi^2+(\beta+it)^2]}, \quad t\in\mathbb{R},$$

where i is the imaginary unit, *i.e.*, satisfying $i^2=-1$.

In addition, as its name suggests, the moment generating function allows moments of different orders to be derived. The moment of order 1, *i.e.*, the mean, is calculated in the result below.

Proposition 2.4 *The mean of a random variable X having the AC distribution with parameters $\alpha\in[-1,1]$ and $\beta\in\mathbb{R}$ is given by*

$$\mu_1 = E(X) = \frac{(\beta^2+\pi^2)\beta[\pi^2+(1-\alpha)\beta^2]\coth(\beta)-(1-\alpha)\beta^4-\pi^2(\alpha+2)\beta^2-\pi^4}{\beta(\beta^2+\pi^2)[\pi^2+(1-\alpha)\beta^2]},$$

where $K_{\alpha,\beta}$ is given as Equation (3)*. We recall that* $\coth(x)=\cosh(x)/\sinh(x)$*, with* $\cosh(x)=(e^x+e^{-x})/2$*, for* $x\in\mathbb{R}$*.*

Proof. We can use the relation that exists between the mean and the moment generating function of X determined in Proposition 2.3, which is

$$\mu_1 = E(X) = G'(s)\big|_{s=0},$$

where the derivative is taken with respect to s. After tedious manipulations, we get

$$G'(s) = K_{\alpha,\beta}\left\{-\frac{2[\pi^2+(1-\alpha)(\beta+s)^2]\sinh(\beta+s)}{(\beta+s)^2[\pi^2+(\beta+s)^2]} - \frac{4[\pi^2+(1-\alpha)(\beta+s)^2]\sinh(\beta+s)}{[\pi^2+(\beta+s)^2]^2}\right.$$
$$\left.+\frac{4(1-\alpha)\sinh(\beta+s)}{\pi^2+(\beta+s)^2}+\frac{2[\pi^2+(1-\alpha)(\beta+s)^2]\cosh(\beta+s)}{(\beta+s)[\pi^2+(\beta+s)^2]}\right\}.$$

Therefore, by taking this derivative at $s=0$, we obtain

$$\mu_1 = K_{\alpha,\beta}\frac{2\beta(\beta^2+\pi^2)[\pi^2+(1-\alpha)\beta^2]\cosh(\beta)-2[(1-\alpha)\beta^4+\pi^2(\alpha+2)\beta^2+\pi^4]\sinh(\beta)}{\beta^2(\beta^2+\pi^2)^2}$$
$$= \frac{(\beta^2+\pi^2)\beta[\pi^2+(1-\alpha)\beta^2]\coth(\beta)-(1-\alpha)\beta^4-\pi^2(\alpha+2)\beta^2-\pi^4}{\beta(\beta^2+\pi^2)[\pi^2+(1-\alpha)\beta^2]}.$$

This ends the proof. □

This result is completed with the result below, that determines the moment of order 2 associated with the AC distribution.

Proposition 2.5 *The moment of order* 2 *of a random variable* X *having the AC distribution with parameters* $\alpha \in [-1,1]$ *and* $\beta \in \mathbb{R}$ *is given by*

$$\mu_2 = E(X^2) = \frac{Q_{\alpha,\beta}}{\beta^2(\beta^2+\pi^2)^2[\pi^2+(1-\alpha)\beta^2]},$$

where

$$\begin{aligned} Q_{\alpha,\beta} &= (1-\alpha)\beta^8 + [3\pi^2+2-2(1+\pi^2)\alpha]\beta^6 + \pi^2[(6-\pi^2)\alpha+3(2+\pi^2)]\beta^4 \\ &\quad -2(\beta^2+\pi^2)\beta[(1-\alpha)\beta^4+\pi^2(\alpha+2)\beta^2+\pi^4]\coth(\beta)+\pi^4(6+\pi^2)\beta^2+2\pi^6. \end{aligned}$$

Proof. We can use the relation that exists between the moment of order 2 and the moment generating function of X determined in Proposition 2.3, which is

$$\mu_2 = E(X^2) = G''(s)\big|_{s=0}.$$

After tedious developments, we obtained the stated formula. For the sake of space, the details are omitted. □

These moments are thus expressible, but remain of a certain mathematical complexity. From a random variable X having the AC distribution, we can also derive the moment measures below.

- The variance of X is defined by $V = E(X^2) - [E(X)]^2 = \mu_2 - (\mu_1)^2$. It measures the dispersion of X around its mean. From it, we can also derive the standard deviation of X defined by $\sigma = \sqrt{V}$.
- The moment skewness is defined by

$$SK = E\left[\left(\frac{X-\mu_1}{\sigma}\right)^3\right].$$

  It measures the asymmetry of the distribution. A positive value for SK indicates a longer tail on the right, while a negative value indicates a longer tail on the left. This measure therefore helps to understand the direction and degree of asymmetry.
- The moment kurtosis is determined by

$$KU = E\left[\left(\frac{X-\mu_1}{\sigma}\right)^4\right].$$

  It measures the "tailedness" of the distribution. A high value for KU indicates heavy tails, while a low value indicates light tails. This measure therefore helps to assess the extreme nature of the deviations from the mean.

Obviously, SK and KU cannot be expressed simply. So we provide numerical work to evaluate them.

In Table 2, we determine the values of these measures for selected values of α and β (the same ones considered in Figure 1).

Table 2. Some values of the moment measures associated with the AC distribution for selected values of the parameters α and β.

	μ_1	V	SK	KU
$\alpha=-1,\beta=-10$	−0.909	0.008	1.749	7.165
$\alpha=0.5,\beta=1$	0.225	0.213	−0.362	2.431
$\alpha=-0.3,\beta=-0.7$	−0.264	0.346	0.615	2.087
$\alpha=-0.7,\beta=3$	0.759	0.067	−2.781	14.035
$\alpha=1,\beta=-1$	−0.129	0.126	0.212	2.485

We see that SK can be positive, close to 0 or negative, and that KU can be small or large, illustrating the adaptability of the AC distribution to different statistical scenarios.

We end this section with a secondary result on the expectation of some transformations of a random variable X having the AC distribution.

Proposition 2.6 *Let X be a random variable having the AC distribution with parameters $\alpha\in[-1,1]$ and $\beta\in\mathbb{R}$. Then the moment results below hold.*

1. *We have*

$$E\left[\frac{1}{1+\alpha\cos(\pi X)}\right]=\begin{cases}\dfrac{\pi^2+\beta^2}{\pi^2+(1-\alpha)\beta^2} & \text{if } \beta\neq 0,\\ 1 & \text{if } \beta=0.\end{cases}.$$

2. *We have*

$$E(e^{-\beta X})=2K_{\alpha,\beta}.$$

3. *We have*

$$E\left[\frac{e^{-\beta X}}{1+\alpha\cos(\pi X)}\right]=2K_{\alpha,\beta}.$$

Proof. The proof is mainly based on the law of the unconscious statistician.

1. We have

$$E\left[\frac{1}{1+\alpha\cos(\pi X)}\right]=\int_{-\infty}^{+\infty}\frac{1}{1+\alpha\cos(\pi x)}f(x)dx$$
$$=\int_{-1}^{1}\frac{1}{1+\alpha\cos(\pi x)}K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{\beta x}dx$$
$$=K_{\alpha,\beta}\int_{-1}^{1}e^{\beta x}dx=\begin{cases}\dfrac{\pi^2+\beta^2}{\pi^2+(1-\alpha)\beta^2} & \text{if } \beta\neq 0,\\ 1 & \text{if } \beta=0.\end{cases}$$

2. We have

$$E(e^{-\beta X}) = \int_{-\infty}^{+\infty} e^{-\beta x} f(x)dx = \int_{-1}^{1} e^{-\beta x} K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{\beta x}dx$$
$$= K_{\alpha,\beta}\int_{-1}^{1}[1+\alpha\cos(\pi x)]dx = K_{\alpha,\beta}\left[2+\frac{\alpha}{\pi}(\sin(\pi)-\sin(-\pi))\right] = 2K_{\alpha,\beta}.$$

3. We have

$$E\left[\frac{e^{-\beta X}}{1+\alpha\cos(\pi X)}\right] = \int_{-\infty}^{+\infty}\frac{e^{-\beta x}}{1+\alpha\cos(\pi x)}f(x)dx$$
$$= \int_{-1}^{1}\frac{e^{-\beta x}}{1+\alpha\cos(\pi x)}K_{\alpha,\beta}[1+\alpha\cos(\pi x)]e^{\beta x}dx = K_{\alpha,\beta}\int_{-1}^{1}dx = 2K_{\alpha,\beta}.$$

The desired formulas are obtained. □

The rest of the article is devoted to the practicality of AC distribution.

3. Applications

Some examples of statistical applications of the AC distribution are now presented.

3.1. Parametric estimation

In practice, when the AC distribution may be appropriate for the analysis of some data, the parameters α and β remain unknown. We therefore need to estimate them as best we can from the data. A popular method for doing this is the maximum likelihood (ML) estimation. Let us briefly introduce it. Let n be the number of data and $x_1,\ldots,x_n$ be the data that we know to be contained in $[-1,1]$. Then we define the likelihood function of the AC distribution as $L_{\alpha,\beta} = \prod_{i=1}^{n} f(x_i)$ (where $f(x)$ is the corresponding PDF) and its logarithmic version as $\ell_{\alpha,\beta} = \ln(L_{\alpha,\beta}) = \sum_{i=1}^{n}\ln[f(x_i)]$. Then the ML estimates (MLEs) of α and β, say $\hat{\alpha}$ and $\hat{\beta}$, are obtained as

$$(\hat{\alpha},\hat{\beta}) = \text{argmax}_{(\alpha,\beta)\in[-1,1]\times\mathbb{R}}\ell_{\alpha,\beta}.$$

These estimates can be determined in the R software using the function nlminb. Based on the MLEs, we can estimate the PDF of the AC distribution using the following function:

$$\hat{f}(x) = f(x)|_{(\alpha,\beta)=(\hat{\alpha},\hat{\beta})} = K_{\hat{\alpha},\hat{\beta}}[1+\hat{\alpha}\cos(\pi x)]e^{\hat{\beta}x}, \quad x\in[-1,1].$$

Conceptually, this function must efficiently fit the shape of the normalised histogram of the data. The accuracy of MLEs and $\hat{f}(x)$ is well documented in the literature, even for n of moderate value. More details on this aspect can be found in [16].

Two different criteria can be associated with this estimation: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). In terms of formulas, by taking into account the setting of the AC distribution, the AIC is defined by

$$AIC = 2k - 2\hat{\ell},$$

where k denotes the number of parameters, *i.e.*, $k = 2$, and $\hat{\ell} = \ell_{\hat{\alpha},\hat{\beta}}$, and the BIC is calculated as

$$BIC = k\ln(n) - 2\hat{\ell}.$$

The AIC assesses the quality of the distribution by balancing the goodness of fit and complexity of this distribution. It therefore favours distributions with lower AIC values. The BIC, on the other hand, also assesses distribution fit and complexity, but imposes a stronger penalty on complex distributions. Lower BIC values indicate better distributions. We again refer to [16].

Before presenting our applications, we should mention that any data can be normalized between -1 and 1. For example, if $y_1,\ldots,y_n$ denote n arbitrary data, for any $i = 1,\ldots,n$, we can consider

$$x_i = 2\frac{y_i - \min(y_1,\ldots,y_n)}{\max(y_1,\ldots,y_n) - \min(y_1,\ldots,y_n)} - 1, \tag{6}$$

in such a way that $x_i \in [-1,1]$. Then $x_1,\ldots,x_n$ are data contained in $[-1,1]$ derived from $y_1,\ldots,y_n$.

3.2. Data fitting scenarios

We now illustrate the interest of combining the AC distribution with the ML estimation to fit some simulated data sets in different credible but hypothetical scenarios. Famous real data sets of air quality measurements in New York and seismic events near Fiji are used in the last two scenarios.

Note that due to the lack of asymmetric distributions with support $[-1,1]$, the AC distribution being one of the rare of its trigonometric type, we do not make any comparison with other distributions, but we give the tools to do so in the future; the AIC and BIC are given in all data fitting scenarios.

Scenario 1: We aim to analyze the performance of 13 different stocks. To do this, we use a specific metric that measures the stock's performance relative to the market. The measures are standardized to vary between -1, indicating underperformance, and 1, indicating strong outperformance. The range of $[-1,1]$ is to facilitate comparison between different stocks.The data are: 0.82, 0.45, 0.67, 0.91, -0.23, 0.74, -0.55, 0.63, -0.82, 0.29, 0.97, -0.14, 0.50.

Figure 8 presents the violin plot of these data. We recall that a violin plot is a statistical graph that combines aspects of a box plot and a kernel density plot to visualize the distribution, underlying PDF and summary statistics of a data set. It shows the median, quartiles and potential outliers, while also showing the full distribution of the data, often through a mirrored density plot.

From this figure, we can see that the data are clearly not symmetric and there are no outliers. The AC distribution is a potential candidate to fit them.

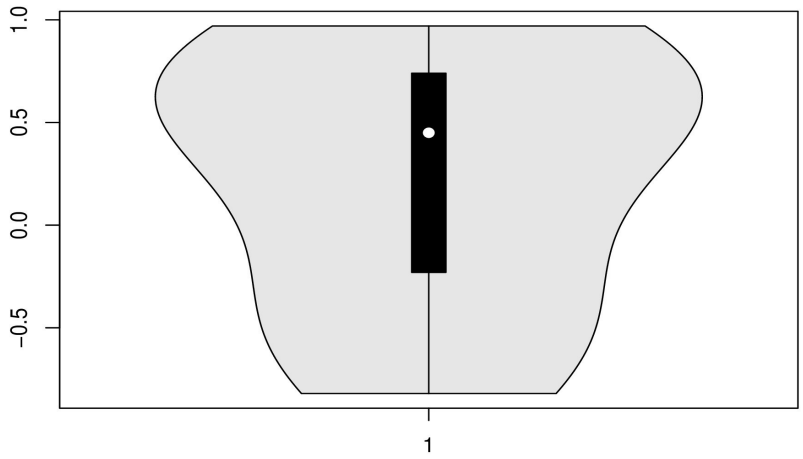

Figure 8. Violin plot of the data in Scenario 1.

After performing the ML estimation to the AC distribution, we obtain $\hat{\alpha} = -0.4433735$ and $\hat{\beta} = 0.5433225$, from which we derive the estimated PDF $\hat{f}(x)$. Figure 9 shows the normalized histogram of the data (with respect to the feature of a PDF) and this estimated PDF.

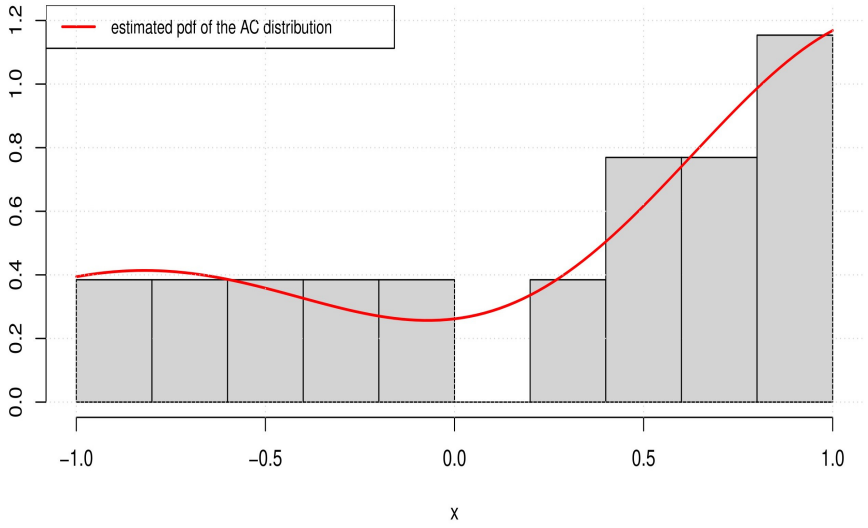

Figure 9. Histogram of the data and estimated PDF of the AC distribution in Scenario 1.

This histogram is not symmetric, but the estimated PDF fits it well. In particular, the rising bars are well captured. This demonstrates the ability of the AC distribution to fit a wide range of data in $[-1, 1]$ beyond the symmetric feature, unlike the cosine distribution. We also find that $AIC = 18.97403$ and $BIC = 20.10393$, which may be useful for further model comparisons for the fit of other distributions based on these data.

Scenario 2: We are looking at normalized data from a study of temperature changes over a 43-year period. Each value therefore corresponds to the normalized change in temperature, ranging from -1 for a significant decrease to 1 for a significant increase, for a given year relative to the baseline temperature average.

The data are: 0.23, −0.45, 0.67, −0.81, 0.92, −0.13, 0.55, −0.72, 0.89, −0.34, 0.78, −0.56, 0.41, −0.98, 0.62, −0.25, 0.87, −0.67, 0.73, −0.84, 0.29, −0.91, 0.48, −0.69,

0.81, −0.52, 0.36, −0.77, 0.93, −0.18, 0.64, −0.63, 0.79, −0.47, 0.58, −0.88, 0.97, −0.09, 0.71,−0.71, 0.85, −0.39, 0.53.

Figure 10 presents the violin plot of these data.

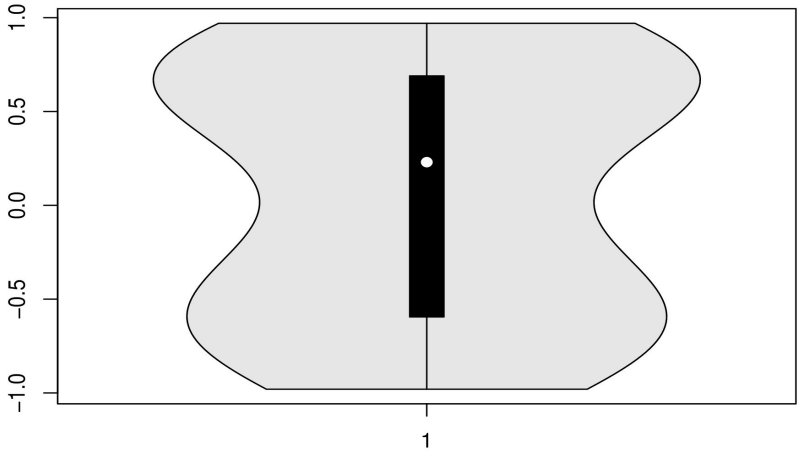

Figure 10. Violin plot of the data in Scenario 2.

From this figure, we can see that the data are of bimodal nature, and clearly not symmetric. There are no outliers. The AC distribution is thus an option to fit them.

Applying the ML estimation to the AC distribution, we obtain $\hat{\alpha} = -0.6529136$ and $\hat{\beta} = 0.1311314$, from which we derive the estimated PDF $\hat{f}(x)$. Figure 11 shows the normalized histogram of the data and this estimated PDF.

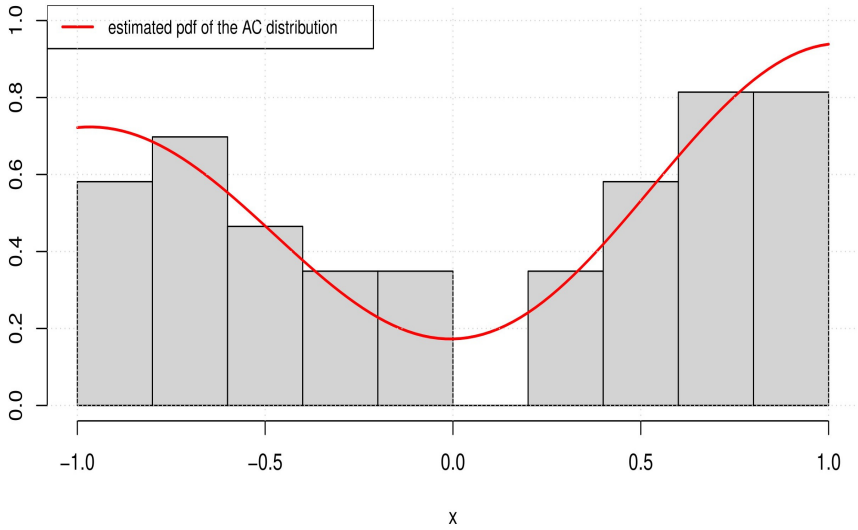

Figure 11. Histogram of the data and estimated PDF of the AC distribution in Scenario 2.

We see that the estimated PDF fits the normalized histogram well. In particular, it captures the bimodal nature of the data. We also find that $AIC = 54.38732$ and $BIC = 57.90972$, which can be useful for further comparisons.

Scenario 3: We consider measurements of a physical quantity taken over 57 time points. Each value is normalized to the range $[-1, 1]$, where a value around 0 is one of the data points indicating a relatively low magnitude of the measured quantity. In this setting, the variations in the data set reflect the natural fluctuations and dynamics of the physical

system being observed.

The data are: 0.87, −0.92, 0.63, −0.15, 0.05, 0.33, −0.76, 0.41, −0.28, 0.11, −0.03, −0.55, 0.72, −0.81, 0.95, 0.05, −0.47, 0.26, −0.69, 0.58, −0.34, 0.19, −0.22, 0.07, −0.91, 0.81, −0.58, 0.69, −0.26, 0.47, −0.05, 0.34, −0.19, 0.22, −0.07, 0.91, −0.81, 0.58, −0.69, 0.26, −0.47, 0.05, −0.34, 0.19, −0.22, 0.07, −0.91, 0.81, −0.58, 0.69, −0.26, 0.47, −0.05, 0.34, −0.19, 0.22, −0.07.

Figure 12 displays the violin plot of these data.

This figure shows relative symmetry in the data and there are no outliers. The AC distribution may be close to symmetric, it is a potential candidate to fit it.

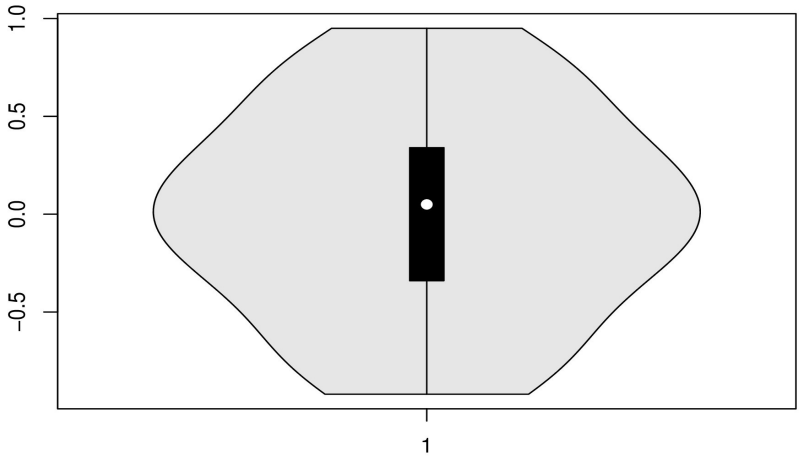

Figure 12. Violin plot of the data in Scenario 3.

Using the ML estimation for the AC distribution, we obtain $\hat{\alpha} = 0.33504720$ and $\hat{\beta} = 0.03106532$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Note that the small value of $\hat{\beta}$ makes the AC distribution close to the cosine distribution. Figure 13 shows the normalized histogram of the data and this estimated PDF.

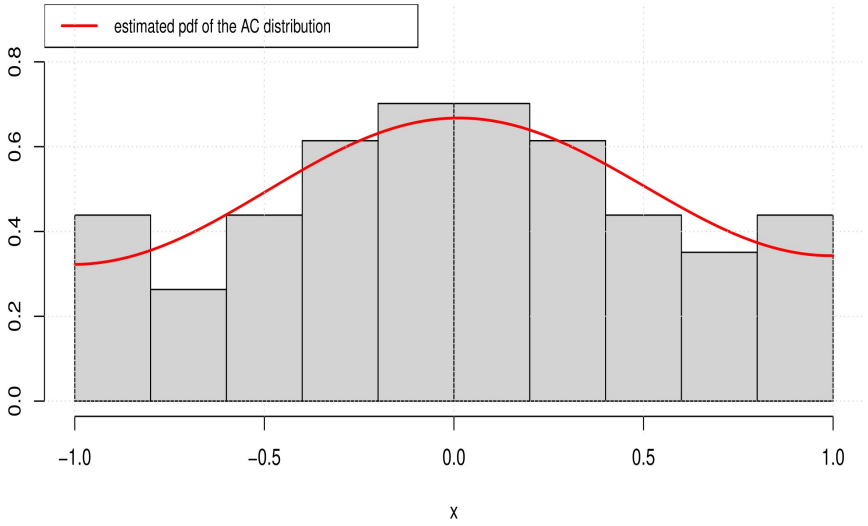

Figure 13. Histogram of the data and estimated PDF of the AC distribution in Scenario 3.

The normalized histogram is well fitted by the estimated PDF. The main symmetry is well captured, and a slight asymmetry to the left is also visible. We also find that

$AIC = 79.51153$ and $BIC = 83.59763$, which can be useful for further comparisons.

Scenario 4: We have a series of measurements taken over time from a sensor array monitoring environmental conditions in a laboratory. There are 49 values ranging from -1 to 1, with some centered around 0 due to the stability of the environment, but with a slight asymmetry due to occasional fluctuations caused by external factors.

The data are: 0.1, -0.5, 0.8, -0.2, -0.3, 0.6, -0.9, 0.4, -0.7, 0.3, 0.2, -0.4, 0.5, -0.1, -0.6, 0.7, -0.8, 0.9, -0.05, -0.15, 0.25, -0.35, 0.45, -0.55, 0.65, -0.75, 0.85, -0.95, 0.01, -0.02, 0.03, -0.04, 0.05, -0.06, 0.07, -0.08, 0.09, -0.11, 0.12, -0.13, 0.14, -0.16, 0.17, -0.18, 0.19, -0.21, 0.22, -0.23, 0.24.

Figure 14 displays the violin plot of these data.

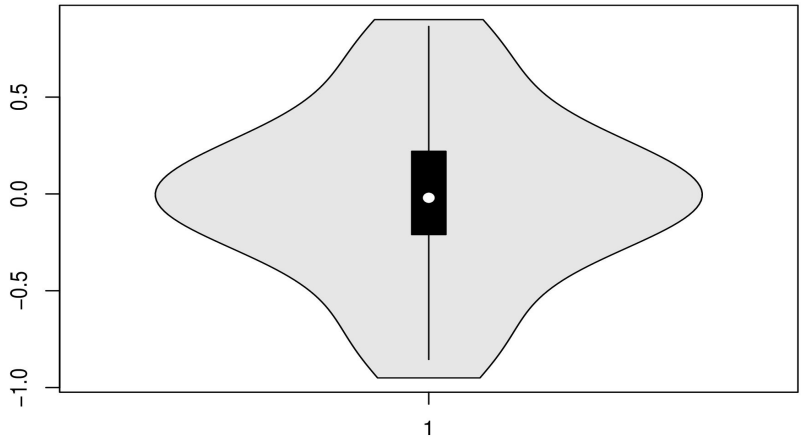

Figure 14. Violin plot of the data in Scenario 4.

This figure shows a slight asymmetry in the data. There are no outliers. The AC distribution can therefore be considered as a statistical model for fitting them.

The ML estimation applied to the AC distribution gives $\hat{\alpha} = 0.64131018$ and $\hat{\beta} = -0.04917237$, and the estimated PDF $\hat{f}(x)$ follows. Note that the small value of $\hat{\beta}$ makes the AC distribution close to the cosine distribution, but with a slightly asymmetric feature. Figure 15 shows the normalized histogram of the data and this estimated PDF.

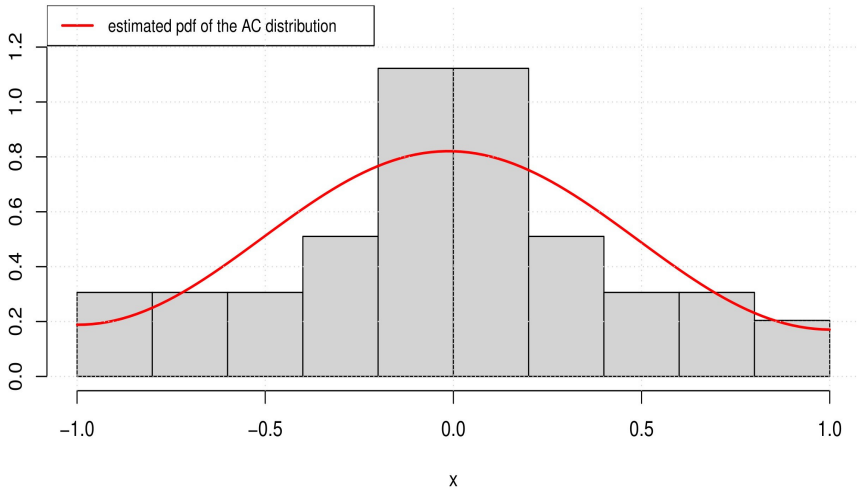

Figure 15. Histogram of the data and estimated PDF of the AC distribution in Scenario 4.

We see that the estimated PDF fits the main form of the normalized histogram, but it slightly misses the mode around $x = 0$. We also find that $AIC = 59.07445$ and $BIC = 62.85809$, which can be useful for further comparisons.

Scenario 5: The data set under consideration contains 56 values. They represent certain monthly temperature anomalies. Each value reflects how much warmer or colder a particular month was compared to the long-term average. Furthermore, each value is rounded to two decimal places and falls within the range $[-1, 1]$. The data are: 0.15, -0.72, -0.84, -0.98, 0.21, -0.45, -0.63, -0.57, -0.91, 0.03, -0.26, 0.47, -0.39, -0.81, -0.62, -0.73, 0.68, -0.95, -0.14, -0.34, -0.58, 0.29, -0.67, 0.11, -0.92, -0.76, -0.83, -0.53, -0.89, -0.42, 0.56, -0.69, 0.34, -0.24, -0.18, -0.93, -0.79, -0.87, -0.97, 0.08, -0.51, -0.71, -0.66, -0.85, 0.39, -0.96, -0.31, -0.38, -0.59, -0.64, -0.49, 0.16, -0.13, -0.75, 0.45, -0.82.

Figure 16 displays the violin plot of these data.

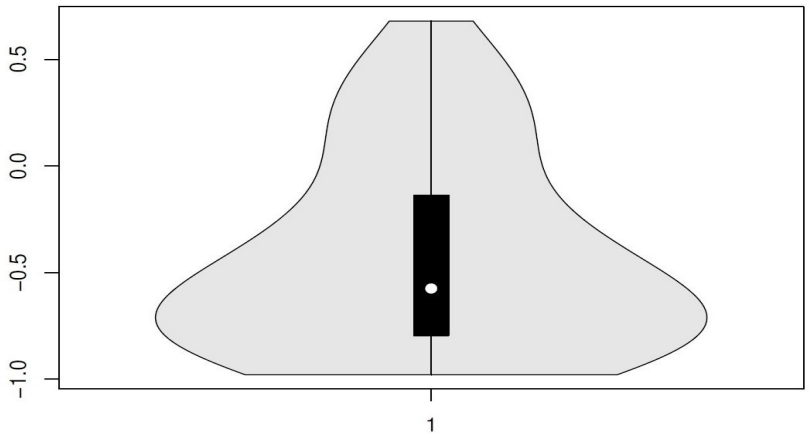

Figure 16. Violin plot of the data in Scenario 5.

This figure shows a strong asymmetry in the data, with a pronounced decreasing pattern and no outliers. The AC distribution is therefore a valuable option for fitting these data. Note that there are more negative values (which indicate colder anomalies) than positive values (which indicate warmer anomalies).

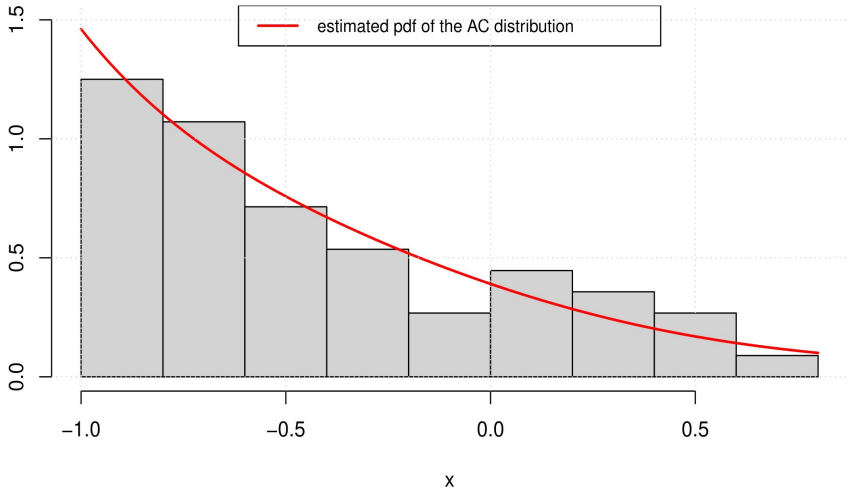

Figure 17. Histogram of the data and estimated PDF of the AC distribution in Scenario 5.

By performing the ML estimation for the AC distribution, we get $\hat{\alpha} = 0.08873247$ and $\hat{\beta} = -1.49784924$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Figure 17 shows the normalized histogram of the data and this estimated PDF.

The decreasing pattern of the normalized histogram is well fitted by the estimated PDF. We also find that $AIC = 50.19003$ and $BIC = 54.24073$. These values may be useful for further distribution comparisons.

Scenario "real life" 6: We now consider at a real data set that includes air quality measurements in New York from May to September 1973. It can be found in [17]. We focus on the variable called wind, whose values are the average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport. The data are contained in the data set entiled airquality of the package datasets included in [18]. The data are airquality [, 3]: 7.4, 8.0, 12.6, 11.5, 14.3, 14.9, 8.6, 13.8, 20.1, 8.6, 6.9, 9.7, 9.2, 10.9, 13.2, 11.5, 12.0, 18.4, 11.5, 9.7, 9.7, 16.6, 9.7, 12.0, 16.6, 14.9, 8.0, 12.0, 14.9, 5.7, 7.4, 8.6, 9.7, 16.1, 9.2, 8.6, 14.3, 9.7, 6.9, 13.8, 11.5, 10.9, 9.2, 8.0, 13.8, 11.5, 14.9, 20.7, 9.2, 11.5, 10.3, 6.3, 1.7, 4.6, 6.3, 8.0, 8.0, 10.3, 11.5, 14.9, 8.0, 4.1, 9.2, 9.2, 10.9, 4.6, 10.9, 5.1, 6.3, 5.7, 7.4, 8.6, 14.3, 14.9, 14.9, 14.3, 6.9, 10.3, 6.3, 5.1, 11.5, 6.9, 9.7, 11.5, 8.6, 8.0, 8.6, 12.0, 7.4, 7.4, 7.4, 9.2, 6.9, 13.8, 7.4, 6.9, 7.4, 4.6, 4.0, 10.3, 8.0, 8.6, 11.5, 11.5, 11.5, 9.7, 11.5, 10.3, 6.3, 7.4, 10.9, 10.3, 15.5, 14.3, 12.6, 9.7, 3.4, 8.0, 5.7, 9.7, 2.3, 6.3, 6.3, 6.9, 5.1, 2.8, 4.6, 7.4, 15.5, 10.9, 10.3, 10.9, 9.7, 14.9, 15.5, 6.3, 10.9, 11.5, 6.9, 13.8, 10.3, 10.3, 8.0, 12.6, 9.2, 10.3, 10.3, 16.6, 6.9, 13.2, 14.3, 8.0, 11.5. We can normalized these data between -1 and 1 using the transformation given in Equation (6) for a better understanding. The normalized data (rounded to two decimals to ease the presentation) are: -0.40, -0.34, 0.15, 0.03, 0.33, 0.39, -0.27, 0.27, 0.94, -0.27, -0.45, -0.16, -0.21, -0.03, 0.21, 0.03, 0.08, 0.76, 0.03, -0.16, -0.16, 0.57, -0.16, 0.08, 0.57, 0.39, -0.34, 0.08, 0.39, -0.58, -0.40, -0.27, -0.16, 0.52, -0.21, -0.27, 0.33, -0.16, -0.45, 0.27, 0.03, -0.03, -0.21, -0.34, 0.27, 0.03, 0.39, 1.00, -0.21, 0.03, -0.09, -0.52, -1.00, -0.69, -0.52, -0.34, -0.34, -0.09, 0.03, 0.39, -0.34, -0.75, -0.21, -0.21, -0.03, -0.69, -0.03, -0.64, -0.52, -0.58, -0.40, -0.27, 0.33, 0.39, 0.39, 0.33, -0.45, -0.09, -0.52, -0.64, 0.03, -0.45, -0.16, 0.03, -0.27, -0.34, -0.27, 0.08, -0.40, -0.40, -0.40, -0.21, -0.45, 0.27, -0.40, -0.45, -0.40, -0.69, -0.76, -0.09, -0.34, -0.27, 0.03, 0.03, 0.03, -0.16, 0.03, -0.09, -0.52, -0.40, -0.03, -0.09, 0.45, 0.33, 0.15, -0.16, -0.82, -0.34, -0.58, -0.16, -0.94, -0.52, -0.52, -0.45, -0.64, -0.88, -0.69, -0.40, 0.45, -0.03, -0.09, -0.03, -0.16, 0.39, 0.45, -0.52, -0.03, 0.03, -0.45, 0.27, -0.09, -0.09, -0.34, 0.15, -0.21, -0.09, -0.09, 0.57, -0.45, 0.21, 0.33, -0.34, 0.03. Mention that the 8 decimal versions of these data are used in the analysis.

Figure 18 displays the violin plot of these data.

This figure shows a slight asymmetry in the data. There are no outliers. The AC distribution can therefore be considered as a statistical model for fitting them.

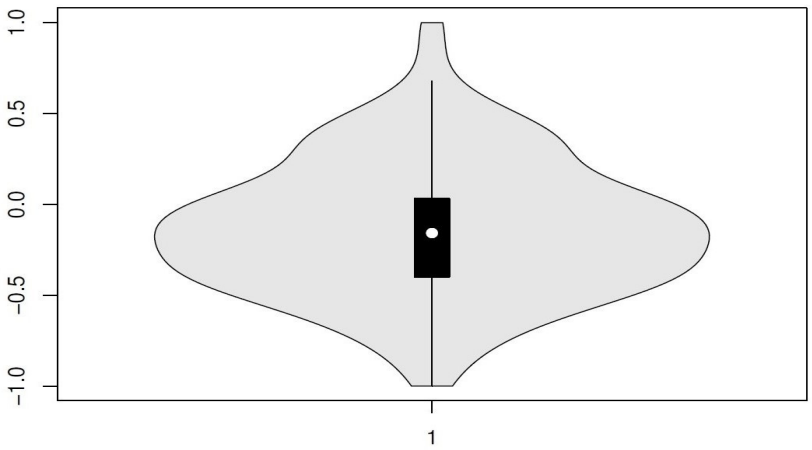

Figure 18. Violin plot of the data in Scenario 6.

By performing the ML estimation for the AC distribution, we obtain $\hat{\alpha} = 0.8976948$ and $\hat{\beta} = -0.8719438$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Figure 19 shows the normalized histogram of the data and this estimated PDF.

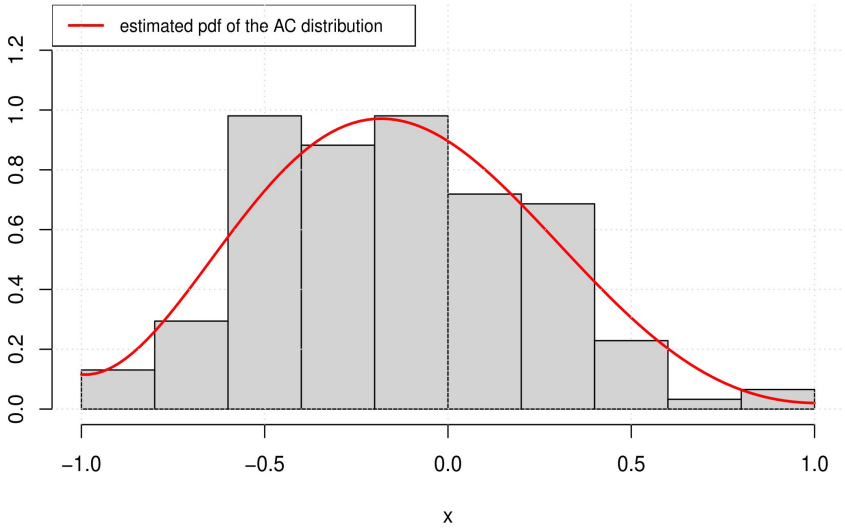

Figure 19. Histogram of the data and estimated PDF of the AC distribution in Scenario "real life" 6.

The bell-shape form of the normalized histogram is well fitted by the estimated PDF. We also find that $AIC = 129.5582$ and $BIC = 135.6191$. As for the previous scenarios, these values may be useful for further distribution comparisons.

Scenario "real life" 7: The data set entiled quakes of the package datasets included in [18] contains the locations of 1000 seismic events (with "MB > 4"). The events have occurred in a cube near Fiji since 1964. The source in the package datasets indicates: "This is one of the Harvard PRIM-H project data sets. They in turn obtained it from Dr. John Woodhouse, Dept. of Geophysics, Harvard University".We consider the following variables:

(i) Latitude of event, corresponding to quakes [, 1],

(ii) Depth (km), corresponding to quakes [, 3],

(iii) Richter magnitude, corresponding to quakes [, 4],

(iv) Number of stations reporting, corresponding to quakes [, 5].

An interesting feature of these variables is that they are of different nature and give rise to different data repartitions (distributions) of the corresponding observations. We therefore show that the AC distribution, with appropriate statistical treatment, is able to adapt to these situations. For each of the variables, the data are normalized between -1 and 1 using the transformation given in Equation (6). The analyses using the AC distribution are given below.

(i) Latitude of event: For the normalized data associated with the event latitude variable, we display the associated violin plot in Figure 20.

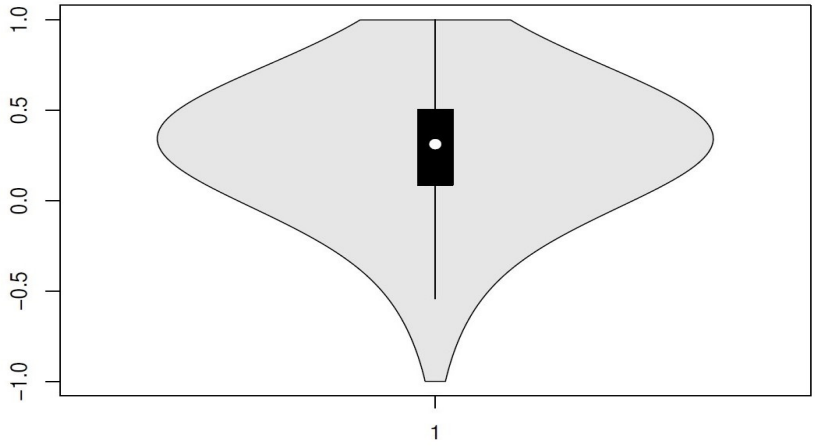

Figure 20. Violin plot of the data in Scenario "real life" 7, variable Latitude of event.

This figure shows a pronounced asymmetry in the data, with no outliers. The AC distribution can therefore be considered as a statistical model to fit the data.

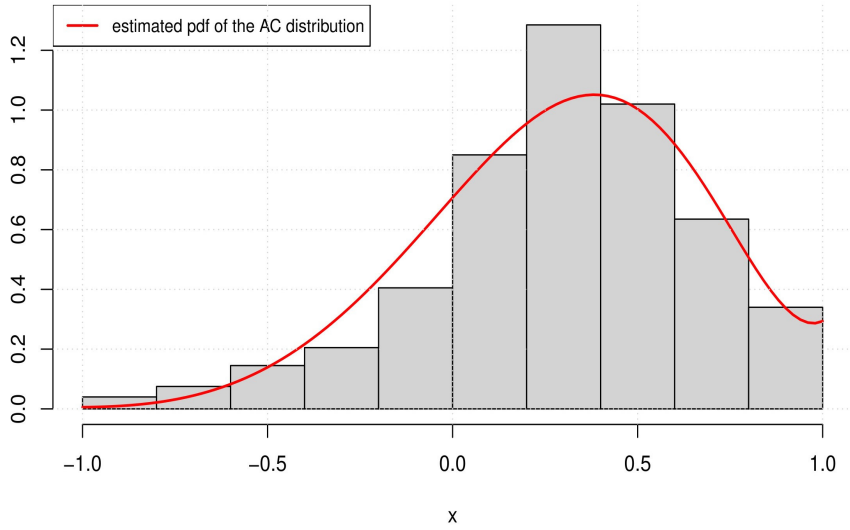

Figure 21. Histogram of the data and estimated PDF of the AC distribution in Scenario "real life" 7, variable Latitude of event.

Applying the ML estimation for the AC distribution, we obtain $\hat{\alpha} = 0.8910731$ and $\hat{\beta} = 1.9753401$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Figure 21 shows the normalized histogram of the data and this estimated PDF.

The left-skewed form of the normalized histogram is well fitted by the estimated PDF. We also find that $AIC = 729.5456$ and $BIC = 739.3611$, which can be used to compare distributions in the future.

(ii) Depth: For the normalized data associated with the depth variable, we show the associated violin plot in Figure 22.

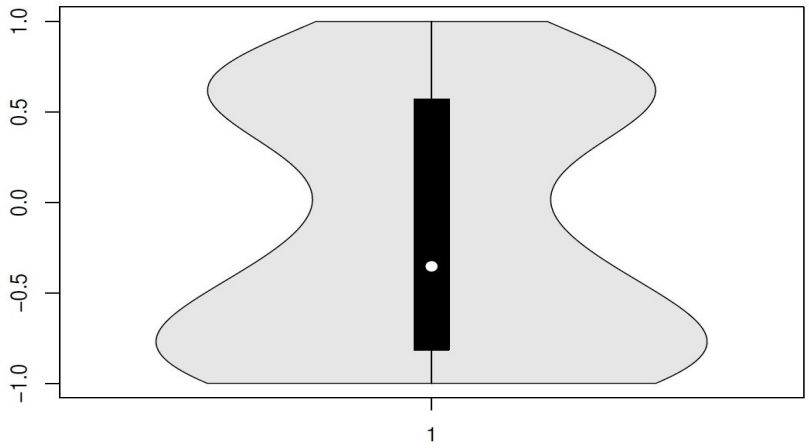

Figure 22. Violin plot of the data in Scenario "real life" 7, variable Depth.

There is a bimodal nature to the data, with no outliers. We can consider using the AC distribution to fit these data.

Performing the ML estimation for the AC distribution, we obtain $\hat{\alpha} = -0.6887218$ and $\hat{\beta} = -0.3256738$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Figure 23 shows the normalized histogram of the data and this estimated PDF.

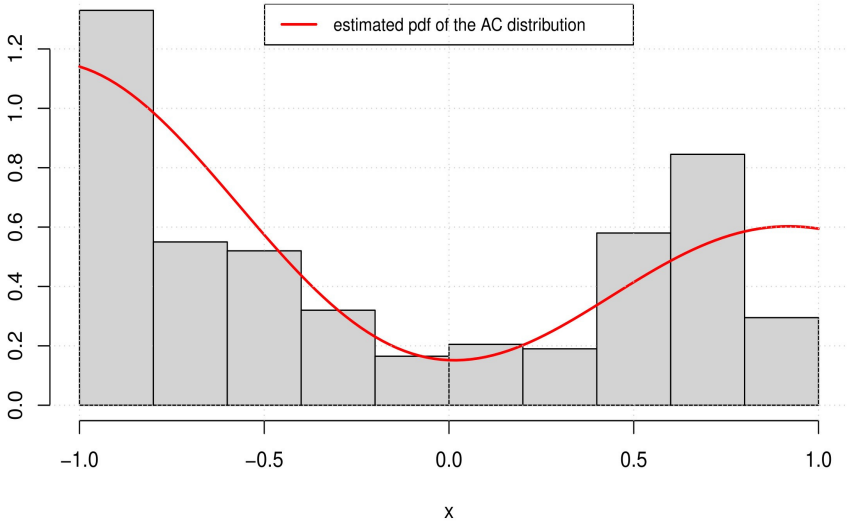

Figure 23. Histogram of the data and estimated PDF of the AC distribution in Scenario "real life" 7, variable Depth.

The bimodal normalized histogram and gap are well fitted by the estimated PDF. We also find that $AIC = 1078.040$ and $BIC = 1087.855$, which can be used to compare distributions in the future.

(iii) Richter magnitude: For the normalized data associated with the Richter magnitude

variable, we display the associated violin plot in Figure 24.

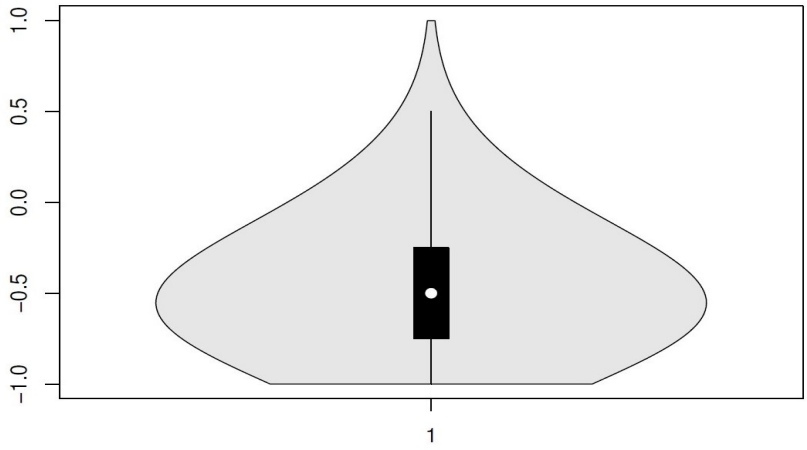

Figure 24. Violin plot of the data in Scenario "real life" 7, variable Richter magnitude.

From this figure, the asymmetry of the data appears. There are no outliers. The AC distribution is therefore designed to fit these data.

Applying the ML estimation for the AC distribution, we obtain $\hat{\alpha} = 0.8459064$ and $\hat{\beta} = -3.4211091$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Figure 25 shows the normalized histogram of the data and this estimated PDF.

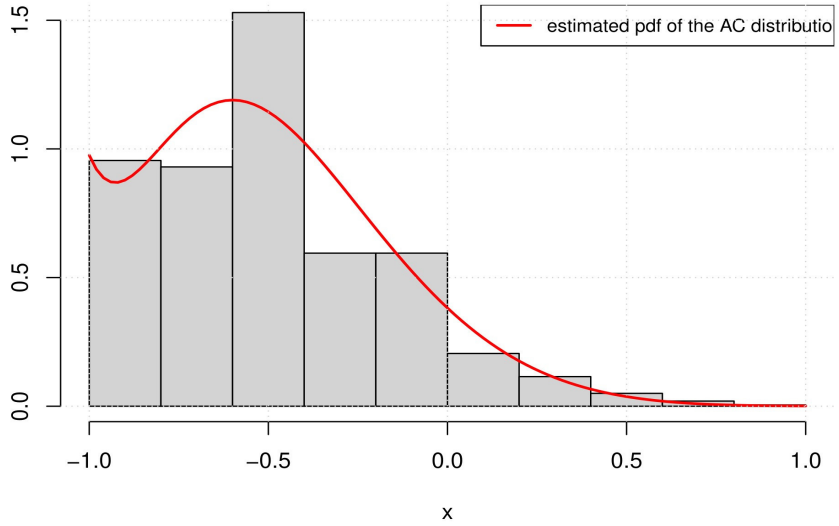

Figure 25. Histogram of the data and estimated PDF of the AC distribution in Scenario "real life" 7, variable Richter magnitude.

The right-skewed form of the normalized histogram is well fitted by the estimated PDF. However, the mode is not fully captured. We also calculate $AIC = 432.5153$ and $BIC = 442.3309$, which can be used to compare distributions in the future.

(iv) Number of stations reporting: For the normalized data associated with the number of stations reporting variable, we present the associated violin plot in Figure 26.

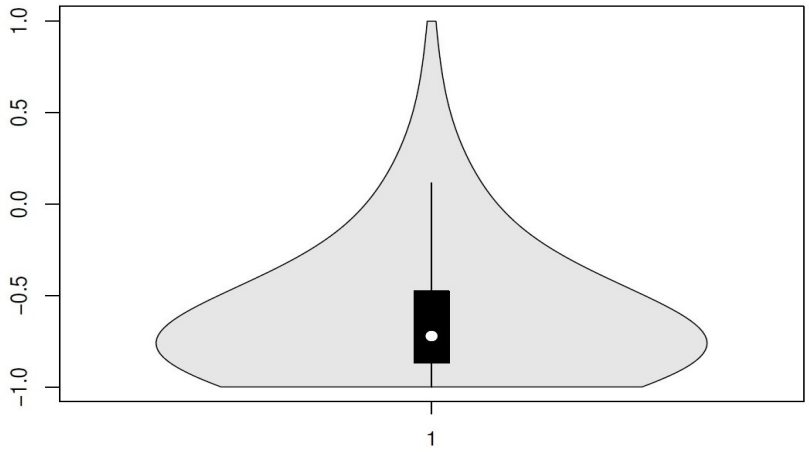

Figure 26. Violin plot of the data in Scenario "real life" 7, variable Number of stations reporting.

From this figure, the decreasing pattern of the data is obseved. There are no outliers. The AC distribution is therefore suitable to fit these data.

Using the ML estimation for the AC distribution, we obtain $\hat{\alpha} = -0.08068575$ and $\hat{\beta} = -2.40712454$. From these estimates, we derive the estimated PDF $\hat{f}(x)$. Figure 27 shows the normalized histogram of the data and this estimated PDF.

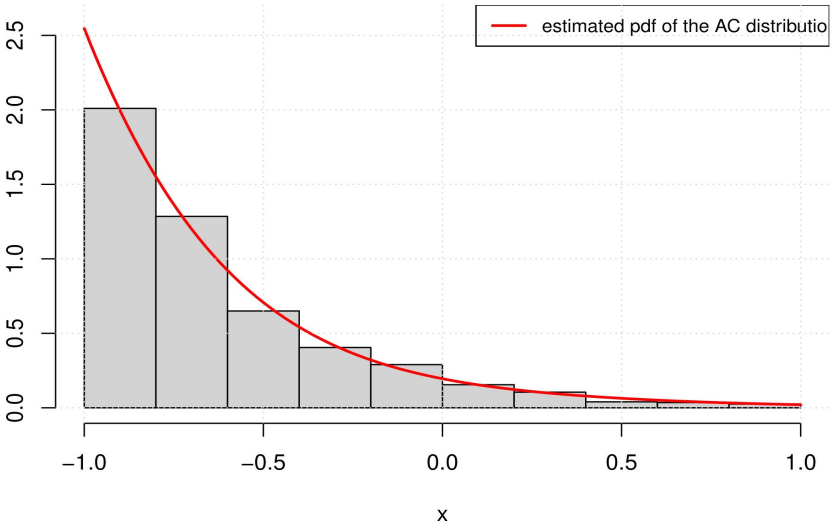

Figure 27. Histogram of the data and estimated PDF of the AC distribution in Scenario "real life" 7, variable Number of stations reporting.

The decreasing form of the normalized histogram is well fitted by the estimated PDF. We also find that $AIC = 76.39319$ and $BIC = 86.20870$, which can be used to compare distributions in the future.

4. Conclusion

In this article, we introduced a modified version of the cosine distribution that is designed to capture asymmetry in data in $[-1, 1]$. It is called the asymmetric cosine (AC) distribution. From a theoretical point of view, we have determined its key functions, highlighting the remarkable flexibility of their shapes, studied the quantiles and carried out a complete moment study. A

statistical part is given on simulated data considered in different credible scenarios, and real data scenarios on air quality and seismic events. It is shown that the AC distribution can be considered as a serious option to fit a wide panel of data with values in $[-1,1]$, regardless of their asymmetry pattern. Multimodality is also covered by the AC distribution.

To the best of our knowledge, this was the first attempt to make the cosine distribution more adaptable to practical scenarios involving asymmetry in the data. Much more can be done on the applied aspects, in particular the application of the AC distribution to fit more real data, its comparison with other comparable asymmetric distributions with support $[-1,1]$ (which is still very rare in the current literature), its consideration in regression-type models, and also its use in modern statistical methods in machine learning, among others.

A possible parallel work can be to use the AC distribution to construct other distributions. For example, based on a random variable X having the AC distribution, we define a new distribution with support $[0,1]$ by considering that of $Y=(1+X)/2$. It is represented by the following modified PDF:

$$f_\star(x)=2f(2x-1)=\begin{cases}2K_{\alpha,\beta}[1-\alpha\cos(2\pi x)]e^{\beta(2x-1)} & \text{if } x\in[0,1],\\ \\ 0 & \text{otherwise.}\end{cases}$$

We believe that it is of real interest because we have transferred the flexibility of the PDF of the AC distribution to the interval $[0,1]$, opening up some modelling options for the analysis of data contained in this interval, such as proportions, frequencies and percentages. For this reason, such distributions have attracted some interest in recent years. See, for example, [19] and [20]. There are perspectives that we have left for future work.

Acknowledgments

The author has received no funds.

Conflicts of Interests

The author declares no conflict of interest.

References

[1] Johnson NL, Kotz S, Balakrishnan N. *Continuous univariate distributions* 2nd ed. New York: Wiley-Interscience, 1995.

[2] Sinha RK. A thought on exotic statistical distributions. *J. Math. Comput. Sci.* 2012, 6(1):49–52.

[3] King M. *Statistics for process control engineers: a practical approach* 1st ed. New York: Wiley, 2017.

[4] Kyurkchiev V, Kyurkchiev N. On the approximation of the step function by raised-cosine and laplace cumulative distribution functions. *Eur. Int. J. Sci. Technol.* 2016, 4(9):75–84.

[5] Rinne H. *Location – scale distributions*: *linear estimation and probability plotting*. Hesse: JLUpub, 2010.

[6] Willink R. *Measurement uncertainty and probability* 1st ed, New York: Cambridge University Press, 2013.

[7] Ahsanullah M, Shakil M. Some characterizations of raised cosine distribution. *J*. *Adv*. *Stat*. *Probab*. 2018, 6(2):42–49.

[8] Watagoda LCRP, Don HSRA, Sanqui JAT. A cosine approximation to the skew normal distribution. *Int*. *Math*. *Forum* 2019, 14(7):253–261.

[9] Ahsanullah M, Shakil M, Kibria BMG. On a generalized raised cosine distribution: some properties, characterizations and applications. *Moroc*. *J*. *Pure Appl*. *Anal*. 2019, 5(1):63–85.

[10] Gradshteyn IS, Ryzhik IM. *Table of integrals*, *series*, *and products* 7th ed. New York: Academic Press, 2007.

[11] Bain LJ, Gan GX. Conditional maxima and inferences for the truncated exponential distribution. *Can*. *J*. *Stat*. 1996, 24(2):251–256.

[12] R Core Team. R: a language and environment for statistical computing. 2016. Available: https://www.R-project.org/ (accessed on 17 May 2024).

[13] Kenney JF, Keeping ES. *Mathematics of statistics part one* 3rd ed. New Jersey: Princeton, 1962, pp.101–102.

[14] Moors JJ. A quantile alternative for kurtosis. *J*. *R*. *Stat*. *Soc*. *D*. 1998, 37(1):25–32.

[15] Bakouch HS, Chesneau C, Leao J. A new lifetime model with a periodic hazard rate and an application. *J*. *Stat*. *Comput*. *Simul*. 2018, 88(11):2048–2065.

[16] Casella G, Berger RL. *Statistical inference* 2nd ed. Pacific Grove: Duxbury Press, 2002.

[17] Chambers JM, Cleveland WS, Kleiner B, Tukey PA. *Graphical methods for data analysis*. Belmont: Wadsworth International Group, 1983.

[18] R Core Team. Package datasets including in R. 2023. Available: https://www.R-project.org/ (accessed on 12 March 2024).

[19] Mazucheli J, Menezes AF, Dey S. Unit-Gompertz distribution with applications. *Statistica* 2019, 79(1):25–43.

[20] Korkmaz MÇ, Altun E, Chesneau C, Yousof HM. On the unit-Chen distribution with associated quantile regression and applications. *Math*. *Slovaca*. 2022, 72(3):765–786.