Article

Kaniadakis's Information Geometry of Compositional Data

Giovanni Pistone [1],* and Muhammad Shoaib [2]

[1] De Castro Statistics, Collegio Carlo Alberto, 10122 Torino, Italy
[2] Department of Mathematics, University of Genoa, 16144 Genova, Italy; muhammad.shoaib@dima.unige.it
* Correspondence: giovanni.pistone@carloalberto.org

Abstract: We propose to use a particular case of Kaniadakis' logarithm for the exploratory analysis of compositional data following the Aitchison approach. The affine information geometry derived from Kaniadakis' logarithm provides a consistent setup for the geometric analysis of compositional data. Moreover, the affine setup suggests a rationale for choosing a specific divergence, which we name the Kaniadakis divergence.

Keywords: Kaniadakis logarithm; information geometry; compositional data; affine displacement; affine statistical bundle; barycenter; Kaniadakis divergence

1. Introduction

This paper describes Kaniadakis' statistics as a methodology in data science. Precisely, we discuss Kaniadakis' formalism for defining an affine structure on the open probability simplex. We present the methods in some generality and use them for the exploratory analysis of compositional data. The illustrating example is a small dataset, and we do not discuss any scaling issues of our methods. However, the dataset has an independent interest in financial risk analysis.

1.1. Why a Geometric Methodology

Kaniadakis' logarithm [1,2] generalises the ordinary logarithm in a way that supports the development of deformed exponential families, deformed statistical divergences, and deformed entropy. Kaniadakis was originally motivated by the applications to non-extensive statistical physics in the sense of [3,4]. In this paper, we present the geometry of the probability simplex as a system of two affine spaces in duality from the perspective of information geometry (IG) [5]. The affine setup was first applied to deformed statistical models in [6].

The systematic use of this formal geometric perspective provides a robust and unified rationale for discussing key descriptive concepts. Defining geometry is much more than providing a topology or a distance. We provide a definition of affine geodesics and a natural duality so that the orthogonal surfaces of the geodesics are well-defined by a specific divergence function. The divergence level sets form a neighbourhood system and, eventually, a topology. In this setup, we define the barycentre, the displacement from the barycentre, and dimensionality reduction. For the standard affine geometry of the probability simplex, see, for example, the tutorial reference [7]. We use a special kind of Kaniadakis' logarithm that appears with a different name in compositional data (CoDa) ([8] Example 4.20).

1.2. CoDa

Compositional data (or CoDa) are the data where all of a (row) vector's (i.e., $[x_1, x_2, \ldots, x_D]$) components are strictly positive real values, can also have zero values, and thus contain solely relative information; the composition is called a D-part composition. Compositional

data are often expressed in closed form and totalled a fixed value, such as 1 for parts per unit or 100 for percentage measurements ([9] Chapter 2).

Compositional data are often found in geosciences and other scientific disciplines, and classification, discrimination, and categorization need to be adapted to the case of CoDa. CoDa analysis is closely related to geosciences and biology, where the data are mostly expressed as proportions or concentrations without mentioning the total size or amount explicitly [10].

Significant advancement has been accomplished during the last thirty to forty years. Recently, the term CoDa analysis has been employed to "Insist on the idea that the study goals or hypotheses, which place more of an emphasis on relative than absolute values, are what ultimately determine composition rather than the data, which may not be pieces of a whole or may not have a fixed sum" [11]. These qualities make CoDa analysis the most powerful tool for applications outside the tradition of hard sciences [12]. Current studies in management, economics, and social sciences have shown in practice the benefits of compositional methods in handling a wide variety of problems, which range from market shares and customer segmentation to tourism, transport systems, financial ratios, and many more (see [11,13–15]).

1.3. CoDa and Systemic Financial Risk

The Center for Risk Management at the University of Lausanne (http://www.crml.ch, accessed on 28 June 2023) provides systemic risk assessments for European financial institutions, which we used in our empirical study using the above Kaniadakis methods. The dataset enables the determination of SRISK country-level values, a market-based systemic risk indicator first proposed in [16,17] and most recently examined in [18].

The characteristics of SRISK are popular in the literature, and SRISK is mainly used to recognize weak institutions and countries with a system-wide impact before a crisis occurs [19] and can help forecast actual sector performance [20].
Most of the previous literature has mainly focused on the absolute values of SRISK. In this work, we focus on implementing the Kaniadakis methods to see the different European countries as a part of compositional data. We developed work started in [12], where they first introduced compositional data analysis to examine the distribution of relative contributions to SRISK connected with key European nations from 2008 to 2021.

Atchison [21] first introduced CoDa analysis. The research conducted by the [12] on financial data used the Atchison methods to examine how European nations contribute to the total amount of systemic risk (SRISK). They find that the distinctive quality of CoDa analysis, especially the Atchison geometry, is very effective in determining the threats of possible instability offered by smaller institutions and nations that might not completely emerge from the scale of their systemic risk.

1.4. Data and Methods

This paper first establishes a novel theoretical framework for compositional data using Kaniadakis' logarithm. Second, we implement the Kaniadakis divergence on the compositional data and calculate the exponential and mixture displacements on compositional data. Next, we calculate the barycenter and deviation. The purpose of calculating the barycenter is to check how far the values of SRISK are from their centre value.

We consider ten European economies (Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Switzerland, and the UK) with annual SRISK measurements collected at the end of December for 2008–2021. Every number is stated in billions of Euros. Like most CoDa method applications, the sample does not cover Europe. Therefore, the ten components that make up our SRISK compositions are just a portion of all those that may be used. CoDa analysis, however, is predicated on the basic notion of sub-compositional coherence, which ensures that a compositional study conducted on a subset of components is consistent with the same analysis performed on the entire composition.

1.5. Kaniadakis' Logarithm

We summarize the particular case of Kaniadakis' logarithm with a purely algebraic form. In the suggestive formalism introduced by [22], the generalised logarithms are associated with the reciprocal derivative function A

$$\log_\kappa(x) = \frac{1}{2}\left(x - \frac{1}{x}\right) = \int_1^x \frac{du}{A(u)}, \quad with \quad A(u) = \frac{2u^2}{1+u^2} \tag{1}$$

$$\exp_\kappa(y) = \log_\kappa^{-1} y = y + \sqrt{1+y^2} = \exp\int_0^y \frac{dv}{\sqrt{1+v^2}}, \quad x^2 - 2xy - 1 = 0, \quad x > 0.$$

Notice that the growth is linear in both directions.

Notice that the above equation reduces any polynomial in y and $x = \exp_\kappa(y)$, for example:

$$\exp_\kappa(y)^2 = 2y\exp_\kappa(y) + 1,$$
$$\exp_\kappa(y)^3 = \exp_\kappa(y)(2y\exp_\kappa(y) + 1) = \dots,$$

and so on. This is an algebraic feature, and this theory is a case of algebraic statistics [23].

The main known properties of the $\kappa-$logarithm and $\kappa-$exponential are

$$\log_\kappa\left(\frac{1}{x}\right) = -\log_\kappa(x), \quad \exp_\kappa(y)\exp_\kappa(-y) = 1$$

$$\frac{d}{dy}\exp_\kappa(y) = A(\exp_\kappa(y)) = \frac{1}{\sqrt{1+y^2}}\exp_\kappa(y)$$

$$\exp_\kappa(x)\exp_\kappa(y) = \exp_\kappa\left(x\sqrt{1+y^2} + y\sqrt{1+x^2}\right) \tag{2}$$

$$\log_\kappa(x_2) - \log_\kappa(x_1) < \frac{1}{A(x_1)}(x_2 - x_1), \quad x_1 \neq x_2 \tag{3}$$

$$\exp_\kappa(y_2) - \exp_\kappa(y_1) > A(\exp_\kappa(y_1))(y_2 - y_1), \quad y_1 \neq y_2 \tag{4}$$

1.6. Kaniadakis' Exponential form of a Positive Probability Function

If the sample space Ω is a finite set, then the probability simplex on Ω is $\mathcal{P}(\Omega)$, and the open probability simplex is $\mathcal{E}(\Omega)$.

For all $p \in \mathcal{E}(\Omega)$, the function $A \circ p$, A as in Equation (1), is strictly positive and provides a positive weight on Ω. It is proportional, but usually not equal, to a probability function,

$$\widetilde{p} \propto A \circ p, \quad namely, \quad \widetilde{p}(x) = \frac{A(p(x))}{\sum_{x\in\Omega} A(p(x))}.$$

We will also write $\widetilde{p} = A \circ p / \overline{A \circ p}$. The mapping $\mathcal{E}(\Omega) \ni p \mapsto \widetilde{p} \in \mathcal{E}(\Omega)$ is called the *escort* mapping; see [22]. See ([24] §3.1) for a discussion of its injectivity and surjectivity. We introduce a notation for the escort expectation, $\widetilde{\mathbb{E}}_p[u] = \sum_{x\in\Omega} u(x)\widetilde{p}(x)$.

For $p, q \in \mathcal{E}(\Omega)$, the *Kaniadakis divergence* can be defined by changes in the usual definition of the logarithm to the Kaniadakis logarithm and the probability function p with the escort $\widetilde{p}$:

$$\begin{aligned}
\widetilde{D}(p|q) &= \widetilde{\mathbb{E}}_p[\log_\kappa(p) - \log_\kappa(q)] \\
&= \sum_{x\in\Omega}(\log_\kappa(p) - \log_\kappa(q)) A \circ p / \overline{A \circ p} \\
&= \sum_{x\in\Omega}(\log_\kappa(p) - \log_\kappa(q))\widetilde{p}(x)
\end{aligned} \tag{5}$$

Clearly, $\widetilde{D}(p|p) = 0$. If $p \neq q$, from the concavity in Equation (3),

$$\widetilde{D}(p|q) = \left(\overline{A \circ p}\right)^{-1} \sum_{x \in \Omega} A(p(x))(\log_\kappa(p(x)) - \log_\kappa(q(x)))$$

$$> \left(\overline{A \circ p}\right)^{-1} \sum_{x \in \Omega} (p(x) - q(x))$$

$$= 0 \, .$$

Fix $p \in \mathcal{E}(\Omega)$. For all $q \in \mathcal{E}(\Omega)$, define

$$s_p(q) = (\log_\kappa(q) - \log_\kappa(p)) + \widetilde{D}(p|q) = (\log_\kappa(q) - \log_\kappa(p)) - \widetilde{\mathbb{E}}_p[(\log_\kappa(q) - \log_\kappa(p))] \, ,$$

then, for $u = s_p(q)$,

$$q = \exp_\kappa\left(u - \widetilde{D}(p|q) + \log_\kappa(p)\right), \quad \widetilde{\mathbb{E}}_p[u] = 0 \, . \tag{6}$$

Conversely, for all $p \in \mathcal{E}(\Omega)$, if u is a random variable such that $\widetilde{\mathbb{E}}_p[u] = 0$, the real function

$$\mathbb{R}_+ \ni \psi \mapsto \sum_{x \in \Omega} \exp_\kappa(u(x) - \psi + \log_\kappa(p(x)))$$

is continuous, goes to 0 as $\psi \to \infty$, and, for $\psi = 0$, takes a value larger than 1 because of Equation (4):

$$\sum_{x \in \Omega} \exp_\kappa(u(x) + \log_\kappa(p(x))) > \sum_{x \in \Omega} \exp_\kappa(\log_\kappa(p(x))) + \sum_{x \in \Omega} A(\exp_\kappa(\log_\kappa(p(x)))u(x)$$

$$= \sum_{x \in \Omega} p(x) + \sum_{x \in \Omega} A(p(x))u(x)$$

$$= 1 \, .$$

In conclusion, there exists a function,

$$K_p \colon S_p = \left\{ u \in L(\Omega) \,\middle|\, \widetilde{\mathbb{E}}_p[u] = 0 \right\} \ni u \mapsto K_p(u) \geq 0 \, ,$$

and $K_p(u) > 0$ provided $u \neq 0$, such that

$$q = \exp_\kappa\left(u - K_p(u) + \log_\kappa(p)\right) \in \mathcal{E}(\Omega) \, .$$

Hence, we have

$$K_p(u) = \widetilde{D}(p|q)$$

and the mapping

$$s_p \colon \mathcal{E}(\Omega) \to S_p$$

is a bijection with inverse

$$e_p \colon S_p \ni u \mapsto \exp_\kappa\left(u - K_p(u) + \log_\kappa(p)\right) \in \mathcal{E}(\Omega) \, . \tag{7}$$

1.7. Properties of the Cumulant Function K_p

Let us compute the derivatives of the function K_p. We use a square bracket notation for the direction

$$dK_p(u)[h] = \lim_{\theta \to 0} \theta^{-1}\left(K_p(u + \theta h) - K_p(u)\right) = \frac{d}{d\theta} K_p(u + \theta h)\Big|_{\theta=0} \, .$$

From Equation (7),

$$
\begin{aligned}
0 &= \frac{d}{d\theta} \sum_{x \in \Omega} \exp_\kappa \big(u(x) + \theta h(x) - K_p(u + \theta h) + \log_\kappa(p(x))\big) \bigg|_{\theta=0} \\
&= \sum_{x \in \Omega} A(\exp_\kappa \big(u(x) + \theta h(x) - K_p(u + \theta h) + \log_\kappa(p(x)))\big)(h(x) - \frac{d}{dt} K_p(u + \theta h)) \bigg|_{\theta=0} \\
&= \sum_{x \in \Omega} A(\exp_\kappa \big(u(x) - K_p(u) + \log_\kappa(p(x)))\big)(h(x) - dK_p(u)[h]) \\
&= \sum_{x \in \Omega} A(e_p(u))(h - dK_p(u)[h]) \ .
\end{aligned}
$$

It follows that, for each $p \in \mathcal{E}(\Omega)$ and $u, h \in S_p$, it holds

$$
dK_p(u)[h] = \widetilde{\mathbb{E}}_q[h] \ , \tag{8}
$$

where $q = e_p(u)$; see Equation (7).

If the curve $t \mapsto q(t)$ has constant divergence, that is, $\widetilde{D}(p|q(t)) = \widetilde{D}(p|q(0))$, derivation provides

$$
0 = \frac{d}{dt} \widetilde{D}(p|q(t)) = \frac{d}{dt} K_p(u(t)) = dK_p(u(t))[\dot{u}(t)] = \widetilde{\mathbb{E}}_{q(t)}[\dot{u}(t)] \ .
$$

Notice that $\widetilde{\mathbb{E}}_{q(0)}[u(t)] = 0$, but this does not imply $\widetilde{\mathbb{E}}_{q(t)}[\dot{u}(t)]$ unless the previous conditions hold true.

1.8. Bibliographical Notes

Similarly, $d^2 K_p$ and the convex conjugate of K_p can be computed. See below for the duality and see also [7,25]. Kaniadakis logarithm and exponential were first introduced in [26,27]. The application to IG used here appeared in [6,24,28]. These papers discuss both the finite state space and the general state space.

2. Affine Space

The Kaniadakis non-parametric affine geometry of the open probability simplex is a variation of the standard case [7]. The main difference is the substitution of the expectation with the escort expectation.

2.1. Statistical Bundle

The statistical bundle is an expression of the tangent space of $\mathcal{E}(\Omega)$ as a dually flat affine statistical manifold in the sense of [5]. The statistical bundle $S\mathcal{E}(\Omega)$ and each fiber $S_q\mathcal{E}(\Omega)$ are defined by

$$
S\mathcal{E}(\Omega) = \left\{ (q, v) \,\middle|\, q \in \mathcal{E}(\Omega), \widetilde{\mathbb{E}}_q[v] = 0 \right\} , \tag{9}
$$

$$
S_q\mathcal{E}(\Omega) = \left\{ v \in L(\Omega) \,\middle|\, \widetilde{\mathbb{E}}_q[u] = 0 \right\} , \quad q \in \mathcal{E}(\Omega) \ . \tag{10}
$$

In our setup, each fibre is a finite-dimensional vector space and can be identified with its dual. However, it is convenient to distinguish the two statistical bundles. The previous one is called *exponential* statistical bundle, while the *mixture* statistical bundle is

$$
{}^*S\mathcal{E}(\Omega) = \left\{ (q, v) \,\middle|\, q \in \mathcal{E}(\Omega), \widetilde{\mathbb{E}}_q[v] = 0 \right\} , \tag{11}
$$

$$
{}^*S_q\mathcal{E}(\Omega) = \left\{ v \in L(\Omega) \,\middle|\, \widetilde{\mathbb{E}}_q[u] = 0 \right\} , \quad q \in \mathcal{E}(\Omega) \ . \tag{12}
$$

For each couple $p, q \in \mathcal{E}(\Omega)$, the mapping

$$
{}^e\mathbb{U}_p^q \colon S_p\mathcal{E}(\Omega) \ni v \mapsto v - \widetilde{\mathbb{E}}_q[v] \in S_q\mathcal{E}(\Omega) \tag{13}
$$

is a bijection. The $^{\mathrm{e}}\mathbb{U}_p^p$ is the identity mapping, and

$$^{\mathrm{e}}\mathbb{U}_q^r \, {}^{\mathrm{e}}\mathbb{U}_p^q = {}^{\mathrm{e}}\mathbb{U}_p^r \, .$$

The co-cycle of mappings $({}^{\mathrm{e}}\mathbb{U}_p^q)_{p,q}$ is the *exponential parallel transport* of the exponential statistical bundle.

The mapping defined for all $p \in \mathcal{E}(\Omega)$, $v \in {}^{*}S_p \, \mathcal{E}(\Omega)$, and $w \in S_p \, \mathcal{E}(\Omega)$ by

$$g \colon (p, v, w) \mapsto g_p(v, w) = \langle v, w \rangle_{A \circ p} = \sum_{x \in \Omega} A(p(x)) u(x) v(x) = \overline{A \circ p} \, \widetilde{\mathbb{E}}_p[vw]$$

provides a duality between the fibres of $S \, \mathcal{E}(\Omega)$ and $^{*}S \, \mathcal{E}(\Omega)$.

The dual of the exponential transport can be computed as follows. For $p, q \in \mathcal{E}(\Omega)$, $v \in {}^{*}S_q \, \mathcal{E}(\Omega)$, and $w \in S_p \, \mathcal{E}(\Omega)$,

$$\begin{aligned}
\left\langle v, {}^{\mathrm{e}}\mathbb{U}_p^q w \right\rangle_{A \circ q} &= \sum_{x \in \Omega} A \circ q \, v(w - \widetilde{\mathbb{E}}_q[w]) \\
&= \sum_{x \in \Omega} A \circ q \, vw - \widetilde{\mathbb{E}}_q[w] \sum A \circ qv \\
&= \sum_{x \in \Omega} A \circ q \, vw \\
&= \sum_{x \in \Omega} A \circ p \left(\frac{A \circ q}{A \circ p} u \right) w \\
&= \left\langle \frac{A \circ q}{A \circ p} u, w \right\rangle_{A \circ p} \, .
\end{aligned}$$

Now, $\frac{A \circ q}{A \circ p} v \in {}^{*}S_p \, \mathcal{E}(\Omega)$; hence, the dual of the exponential transport is the *mixture transport*,

$$^{*}\left({}^{\mathrm{e}}\mathbb{U}_q^p \right) v = {}^{\mathrm{m}}\mathbb{U}_q^p v = \frac{A \circ q}{A \circ p} v \, . \tag{14}$$

2.2. Velocity and Auto-Parallel Curves

The following computation is a version of the original argument about Fisher's score. Let $t \mapsto q(t) \in \mathcal{E}(\Omega)$ be a one-dimensional parametric statistical model, namely a curve in geometric language. We assume the curve is smooth and twice differentiable as a mapping in the vector space $L(\Omega)$. For each random variable $f \in L(\Omega)$,

$$\begin{aligned}
\frac{d}{dt} \mathbb{E}_{q(t)}[f] &= \sum_{x \in \Omega} f(x) \dot{q}(x; t) \\
&= \sum_{x \in \Omega} f(x) \frac{\dot{q}(x; t)}{A(q(x; t))} A(q(x; t)) \\
&= \sum_{x \in \Omega} f(x) \frac{d}{dt} \log_\kappa(q(x; t) A(q(x; t)) \\
&= \sum_{x \in \Omega} A \circ q(t) (f - \mathbb{E}_{\widetilde{q}(t)}[f]) \frac{d}{dt} \log_\kappa(q(t)) \\
&= \left\langle f - \mathbb{E}_{\widetilde{q}(t)}[f], \frac{d}{dt} \log_\kappa(q(t)) \right\rangle_{A \circ q(t)} \\
&= \sum_{x \in \Omega} f(x) \dot{q}(x; t) \\
&= \left\langle f - \mathbb{E}_{\widetilde{q}(t)}[f], \frac{d}{dt} \log_\kappa(q(t)) \right\rangle_{A \circ q(t)}
\end{aligned} \tag{15}$$

The *velocity* of the curve is defined as

$$\overset{\star}{q}(t) = \frac{d}{dt}\log_\kappa(q(t)) = \frac{\dot{q}(t)}{A(q(t))} . \tag{16}$$

We can check that $\overset{\star}{q}(t) \in S_{q(t)}\,\mathcal{E}(\Omega)$ and $(f - \mathbb{E}_{\widetilde{q}(t)}[f]) \in {}^*S_{q(t)}\,\mathcal{E}(\Omega)$. The *Cramer–Rao bound* is

$$\left(\frac{d}{dt}\mathbb{E}_{q(t)}[f]\right)^2 = \overline{A \circ q(t)}^2 \left\langle f - \mathbb{E}_{\widetilde{q}(t)}[f], \frac{d}{dt}\log_\kappa(q(t)) \right\rangle_{\widetilde{q}(t)}$$

$$\leq \overline{A \circ q(t)}^2\,\mathbb{E}_{\widetilde{q}(t)}\left[(f - \mathbb{E}_{\widetilde{q}(t)}[f])^2\right]\mathbb{E}_{\widetilde{q}(t)}\left[\left(\frac{\dot{q}(t)}{A \circ q(t)}\right)^2\right]$$

$$= \sum_{x \in \Omega} A \circ q(t)(f - \widetilde{\mathbb{E}}_q[f])^2 \sum \frac{\dot{q}(t)^2}{A \circ q(t)}$$

$$= \sum_{x \in \Omega} A \circ q(t)(f - \widetilde{\mathbb{E}}_q[f])^2 \sum \frac{\dot{q}(t)^2}{A \circ q(t)} \tag{17}$$

The variation computed with the escort probability function, namely $f - \widetilde{\mathbb{E}}_q[f]$, appears in Equation (15) as a gradient of the expectation $f \mapsto \mathbb{E}_q[f]$.

A curve $t \mapsto q(t)$ is *auto-parallel* for the mixture trasport if

$${}^m\mathbb{U}_{q(t)}^{q(s)}\overset{\star}{q}(t) = \overset{\star}{q}(s) .$$

For $q(0) = q_0$ and $q(1) = q_1$,

$$\frac{\dot{q}(0)}{A \circ q_0} = \frac{A \circ q(t)}{A \circ q_0}\frac{\dot{q}(t)}{A \circ q(t)} = \frac{1}{A \circ q_0}\dot{q}(t) ,$$

so that

$$q(t) = q_0 + t(q_1 - q_0) .$$

Let us compute the auto-parallel curves for the exponential transport,

$${}^e\mathbb{U}_{q(t)}^{q(s)}\overset{\star}{q}(t) = \overset{\star}{q}(s) .$$

For $q(0) = q_0$ and $q(1) = q_1$,

$$\overset{\star}{q}(0) = \overset{\star}{q}(t) - \widetilde{\mathbb{E}}_{q_0}[\overset{\star}{q}(t)] = \frac{d}{dt}\log_\kappa(q(t)) - \frac{1}{A \circ q_0}\sum A \circ q_0\frac{d}{dt}\log_\kappa(q(t)) ,$$

so that, for some function ψ,

$$\log_\kappa(q(t)) = \log_\kappa(q_0) + t\overset{\star}{q}(0) - \psi(t) .$$

Comparing with Equation (7), we have that *the auto-parallel curve for the exponential transport is*

$$q(t) = \exp_\kappa\left(t\overset{\star}{q}(0) - K_{q_0}(t\overset{\star}{q}(0)) + \log_\kappa(q_0)\right) . \tag{18}$$

As observed above,

$$\overset{\star}{q}(0) = \log_\kappa(q_1) - \log_\kappa(q_0) - \widetilde{\mathbb{E}}_{q_0}[\log_\kappa(q_1) - \log_\kappa(q_0)] = \log_\kappa(q_0) + \widetilde{D}(q_0|q_1) .$$

2.3. Surfaces of Constant Divergence

We have observed that an auto-parallel curve starting at $q(0)$ with velocity $\overset{\star}{q}(0)$ has the form of Equation (18). For given extreme points $q_0 = q(0)$ and $q_1 = q(1)$, it holds that

$$q_1 = \exp_\kappa\left(u - K_{q_0}(u) + \log_\kappa(q_0)\right) \quad \text{hence} \quad u = \log_\kappa(q(0)) - \log_\kappa(q(1)) + \widetilde{D}(q_0|q_1)\,,$$

in particular, $\widetilde{\mathbb{E}}_{q_1}[u] = 0$.

The velocity of the auto-parallel curve at q_1 is constant,

$$\frac{d}{dt}\left(tu - K_{q_0}(tu) + \log_\kappa(q_0)\right)\Big|_{t=1} = u - \widetilde{\mathbb{E}}_{q_1}[u] = u.$$

Consider a curve γ starting at $\gamma(0) = q_1$ of the form,

$$t \mapsto \gamma(t) = \exp_\kappa\left(u + v(t) - K_{q_0}(u + v(t)) + \log_\kappa(q_0)\right) \quad \text{with } v(0) = 0$$

and assume a divergence is constant, precisely

$$\widetilde{D}(\gamma(t)|q_0) = \widetilde{D}(\gamma(0)|q_0) = \widetilde{D}(q_1|q_0)\,.$$

It holds

$$\begin{aligned}
\widetilde{\mathbb{E}}_{\gamma(t)}[\log_\kappa(\gamma(t)) - \log_\kappa(q_0)] &= \widetilde{\mathbb{E}}_{\gamma(t)}\left[u + v(t) - K_{q_0}(u + v(t)) + \log_\kappa(q_0) - \log_\kappa(q_0)\right] \\
&= \widetilde{\mathbb{E}}_{\gamma(t)}[u + v(t)] - K_{q_0}(u + v(t)) \\
&= dK_{q(0)}(u + v(t))[u + v(t)] - K_{q_0}(u + v(t))
\end{aligned}$$

is constant so that the derivative is zero. In particular, it is zero at $t = 0$

$$0 = d^2 K_{q_0}[u + v(t)][u + v(t), \dot{v}(t)]\Big|_{t=0} = d^2 K_{q_0}[u, \dot{v}(0)]\,.$$

That is, this surface of equi-divergence is orthogonal to the auto-parallel curves in the sense of the quadratic form d^2K. This is actually the generalization of a well-known result in IG, where the Hessian of the cumulant function is the Fisher's information matrix. See, for example, [5].

2.4. Displacement

The machinery introduced above allows for explicitly defining the affine structure as originally defined by [29]. A textbook on affine geometry is ([30] Ch. 2,3,9). Below, we call the following two (dual) displacements on the statistical bundle. The mixture displacement is

$$\eta_p(q) = \frac{q - p}{A \circ p}\,. \tag{19}$$

The exponential displacement is

$$s_p(q) = (\log_\kappa(q) - \log_\kappa(p)) - \widetilde{\mathbb{E}}_p[(\log_\kappa(q) - \log_\kappa(p))] \tag{20}$$

Both displacements define affine coordinates in the statistical bundle. The easy proofs are the same as in the standard cases [7]. Each displacement defines an atlas of charts on the affine bundle.

The orthogonal surfaces of the affine exponential auto-parallel curves are discussed in the section above. The orthogonal surfaces to the affine mixture auto-parallel curves are easily observed to be associated with the other divergence. In fact, it is the classical result of the duality between the two divergences. See, for example, [5].

The availability of an affine bundle would allow for a coherent and straightforward definition of mechanical concepts such as velocity, acceleration, Lagrangian, and Hamilto-

nian. See [31,32] for the standard case. In the present paper, we develop the application to CoDa, and we stress the notion of affine barycenter and the fact that a system of charts can be observed as a preprocessing of data to be followed by any method adapted to actual vector data.

2.5. Barycenter and Deviation

Let $f_1, \ldots, f_n$ be a sequence of CoDa points with strictly positive components and normalized to one. Each data point is a point in the open probability simplex. The affine coordinates (20) centered at p are

$$s_p(f_1) = (\log_\kappa(f_1) - \log_\kappa(p)) - \widetilde{\mathbb{E}}_p[(\log_\kappa(f_1) - \log_\kappa(p))]$$

$$\vdots$$

$$s_p(f_n) = (\log_\kappa(f_n) - \log_\kappa(p)) - \widetilde{\mathbb{E}}_p[(\log_\kappa(f_n) - \log_\kappa(p))]$$

The *mean value* of the affine coordinates is

$$\overline{s_p} = \frac{1}{n}\sum_{j=1}^{n} s_p(f_j) = \frac{1}{n}\sum_{j=1}^{n}\left(\log_\kappa(f_j) - \widetilde{\mathbb{E}}_p\left[(\log_\kappa(f_j))\right]\right) - \left(\log_\kappa(p) - \widetilde{\mathbb{E}}_p[\log_\kappa(p)]\right) \quad (21)$$

If the mean value computed in the centering q is $\overline{s_q}$, the difference is

$$\overline{s_p} - \overline{s_q} = \log_\kappa(q) - \log_\kappa(p) - \frac{1}{n}\sum_{j=1}^{n}\widetilde{\mathbb{E}}_p\left[\log_\kappa(f_j)\right] + \frac{1}{n}\sum_{j=1}^{n}\widetilde{\mathbb{E}}_q\left[\log_\kappa(f_j)\right].$$

Hence,

$$\overline{s_p} + \log_\kappa(p) = \overline{s_q} + \log_\kappa(q) + \text{constant} .$$

The probability function is the same in both cases. In fact,

$$\exp_\kappa\left(\overline{s_p} - K_p\left(\overline{s_p}\right) + \log_\kappa(p)\right) = \exp_\kappa\left(\overline{s_q} - K_p\left(\overline{s_p}\right) + \log_\kappa(q) + \text{constant}\right)$$
$$= \exp_\kappa\left(\overline{s_q} - K_q\left(\overline{s_q}\right) + \log_\kappa(q)\right)$$

because of the uniqueness of the normalizing constant.

In conclusion, the probability function

$$\overline{f} = \exp_\kappa\left(\overline{s_p} - K_p\left(\overline{s_p}\right) + \log_\kappa(p)\right)$$
$$= \exp_\kappa\left(\overline{s_p} + \widetilde{D}\left(p\middle|\overline{f}\right) + \log_\kappa(p)\right),$$

with $\overline{s_p}$ as Equation (21) *does not depend on the reference p*. It is the *barycentre* of the given data points.

The displacement of each data point f_j from the barycentre $\overline{f}$ is

$$s_{\overline{f}}(f_j) = \overline{s_p} - s_p(f_j) - \widetilde{\mathbb{E}}_{\overline{f}}[\overline{s_p} - s_p(f_j)] = \frac{1}{n}\sum_{k} s_p(f_k) - s_p(f_j) - \widetilde{\mathbb{E}}_{\overline{f}}\left[\frac{1}{n}\sum_{k} s_p(f_k) - s_p(f_j)\right]$$

and the expression of each point f_j in the barycentre $\overline{f}$ is

$$f_j = \exp_\kappa\left(s_{\overline{f}}(f_j) + \widetilde{D}\left(\overline{f}\middle|f_j\right) + \log_\kappa\left(\overline{f}\right)\right).$$

A one-dimensional summary consistent with our formalism of the divergence of each point from the barycentre is the Kaniadakis' divergence $\widetilde{D}\left(\overline{f}\middle|f_j\right)$, which is the normalising constant in the equation above. Another option is the Kaniadakis' divergence $\widetilde{D}\left(f_j\middle|\overline{f}\right)$ that appears in the representation of the barycentre in the data point f_j.

3. Data Analysis

This section will use some geometric concepts derived from Kaniadakis' IG. It should be noted that our formalism is, in principle, affine and does not include any properly defined distance.

3.1. Kaniadakis Divergence

First, we compute the Kaniadakis divergence defined in Equation (5). Each point in (i, j) in Figure 1 is the Kaniadakis divergence of the CoDa point corresponding to the year in the *i*-th row with respect to the CoDa point for the year in the *j*-th column. For example, the Kaniadakis divergence between 2008 and 2009 is $\tilde{D}(2008|2009) = 0.14$. Most values are smaller than one, except when the reference distribution corresponds to 2008 or 2009 for the most recent years. The year 2009 deviates significantly from the other years.

3.2. Mixture Displacement

Equation (19) provides instructions for computing the mixture displacement. From Figure 2, the mixture displacement for Greece and Spain is very high for all the years. The value for Spain in 2009 was less than zero—the only negative value for Spain. On the contrary, all other countries do not have too many high values.

Equation (21) provides the mean value. After determining the mean, we compute the mixture displacement using the mean as a reference. We check that our values abruptly go from -10 to 10. However, the results for Greece and Spain decrease when the mixture displacement from the mean is calculated.
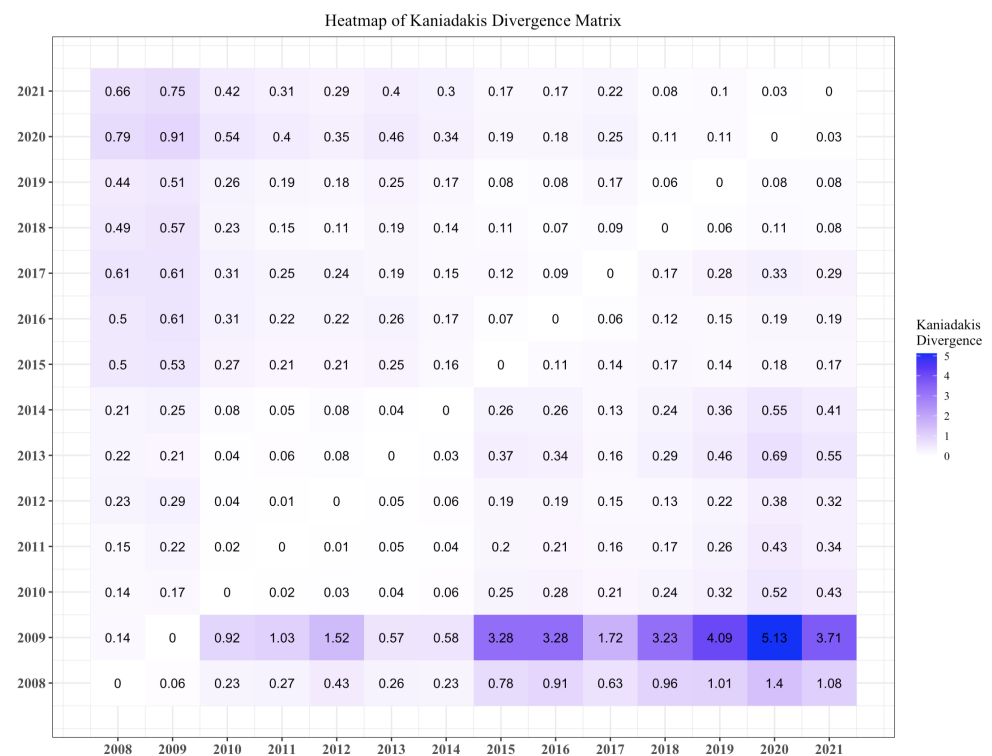
Heatmap of Kaniadakis Divergence Matrix

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
2021	0.66	0.75	0.42	0.31	0.29	0.4	0.3	0.17	0.17	0.22	0.08	0.1	0.03	0
2020	0.79	0.91	0.54	0.4	0.35	0.46	0.34	0.19	0.18	0.25	0.11	0.11	0	0.03
2019	0.44	0.51	0.26	0.19	0.18	0.25	0.17	0.08	0.08	0.17	0.06	0	0.08	0.08
2018	0.49	0.57	0.23	0.15	0.11	0.19	0.14	0.11	0.07	0.09	0	0.06	0.11	0.08
2017	0.61	0.61	0.31	0.25	0.24	0.19	0.15	0.12	0.09	0	0.17	0.28	0.33	0.29
2016	0.5	0.61	0.31	0.22	0.22	0.26	0.17	0.07	0	0.06	0.12	0.15	0.19	0.19
2015	0.5	0.53	0.27	0.21	0.21	0.25	0.16	0	0.11	0.14	0.17	0.14	0.18	0.17
2014	0.21	0.25	0.08	0.05	0.08	0.04	0	0.26	0.26	0.13	0.24	0.36	0.55	0.41
2013	0.22	0.21	0.04	0.06	0.08	0	0.03	0.37	0.34	0.16	0.29	0.46	0.69	0.55
2012	0.23	0.29	0.04	0.01	0	0.05	0.06	0.19	0.19	0.15	0.13	0.22	0.38	0.32
2011	0.15	0.22	0.02	0	0.01	0.05	0.04	0.2	0.21	0.16	0.17	0.26	0.43	0.34
2010	0.14	0.17	0	0.02	0.03	0.04	0.06	0.25	0.28	0.21	0.24	0.32	0.52	0.43
2009	0.14	0	0.92	1.03	1.52	0.57	0.58	3.28	3.28	1.72	3.23	4.09	5.13	3.71
2008	0	0.06	0.23	0.27	0.43	0.26	0.23	0.78	0.91	0.63	0.96	1.01	1.4	1.08

Kaniadakis Divergence: 5, 4, 3, 2, 1, 0

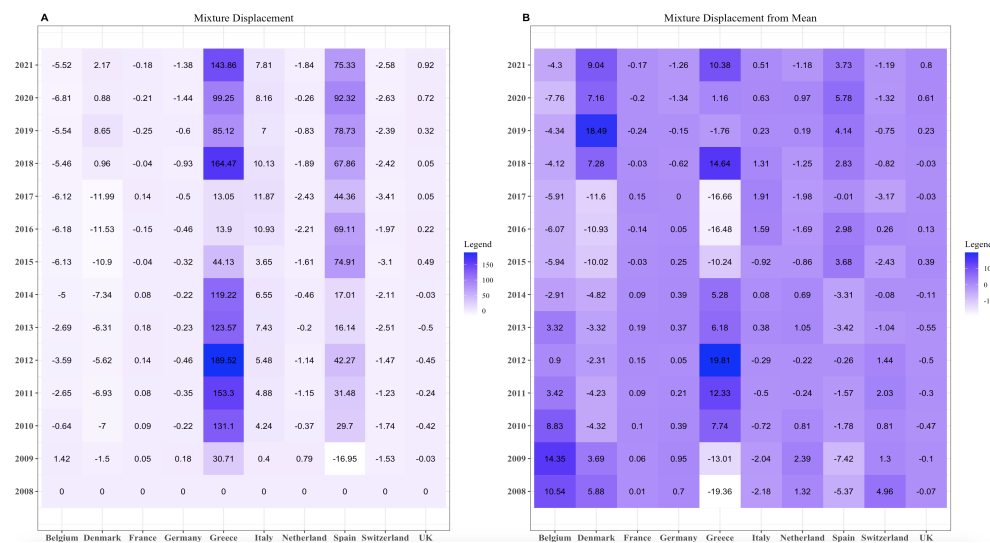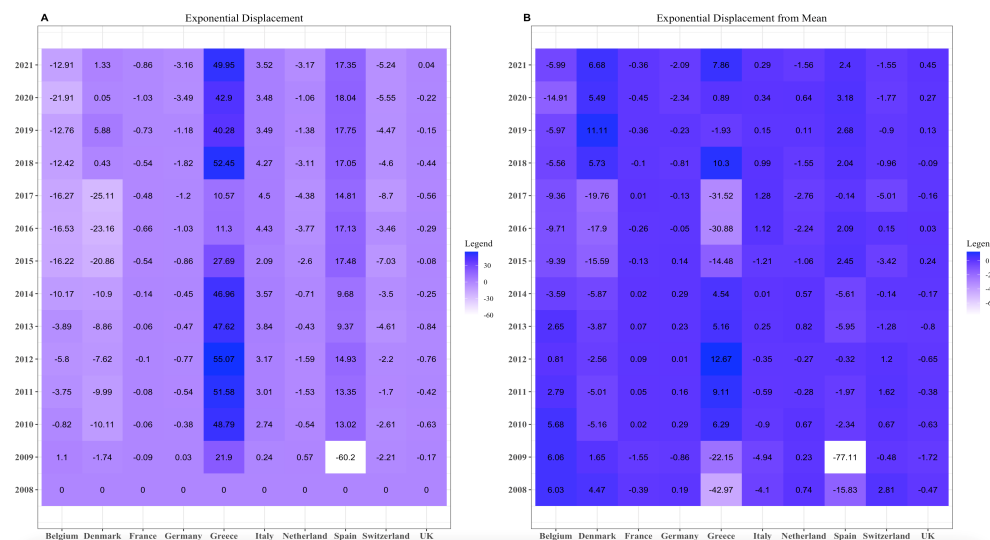**Figure 1.** Kaniadakis divergence on compositional data.

Figure 2. (**A**) Mixture displacement on compositional data by taking 2008 as a reference and (**B**) mixture displacement on compositional data by taking mean as reference.

3.3. Exponential Displacement

As above, Equation (20) returns the exponential displacement. Further, Figure 3 is the empirical result of Equation (20). As for the mixture displacement, we can see that Spain and Greece have higher displacement than other European countries. The value for Spain in the year 2009 is meagre.

If the mean is the reference point, the exponential displacement ranges from 0 to −60. The only significant changes are for the nations of Spain and Greece, where our values for 2009 for Spain decreased by about 18 times, and, for Greece, our values decreased significantly.

Figure 3. (**A**) Exponential displacement on compositional data by taking 2008 as a reference and (**B**) exponential displacement on compositional data by taking mean as a reference.

4. Conclusions and Discussion

In this research, we applied a particular type of divergence, Kaniadakis divergence, to compositional data, aligned with the symmetrised ratio transformation in ([8] Example 4.20). The dataset being examined spans the years 2008 through 2021. First, we built a theoretical framework for Kaniadakis divergence, mixture displacement, and exponential displacement.

Section 1 provided the mathematical framework for determining divergence and displacement, while Section 2 demonstrated how to apply those mathematical algorithms

to compositional data. In the application, we found that Spain and Greece have more fluctuations when compared to the other European countries. The values of the mixture and exponential displacement confirm that Spain and Greece faced some financial crises compared to other countries.

This simple application shows the potential of IG for application to compositional data analysis. We suggest that Kaniadakis' logarithm can reduce the computations for monitoring systemic risk to algebraic computations. The Kaniadakis logarithm, mixture, and exponential displacement on compositional data can be considered to broaden traditional research methods for compositional data analysis.

We would like to add a few words regarding the specific tools and formalism we used here. First, we mimicked one of the possible presentations of non-parametric IG by following the basic dually-flat setup step by step. Another successful presentation of non-parametric IG starts with properly defining the divergences and deriving the geometry; see, for example, [33]. A popular approach, not equivalent to the affine one, defines the geometry of the probability simplex by introducing a metric tensor. As in other geometric theories, one should carefully distinguish between choosing charts and introducing a topology.

In the present approach, we define the charts so that the associated manifold is affine; in this setup, some specific divergences appear as naturally associated with the geometry and the basic statistical notion, namely the pairing between measures and random variables. Everything is applied to simple data manipulation in the spirit of Aitchison's methods.

No claim of optimality is made. The existence of many different but topologically equivalent divergences is only natural in our setup, where the topology actually depends on the geometry and not the other way around. Whenever needed, a choice must be based on some additional assumption. We carefully check the simple, useful operations on data, such that the geodesic connecting two given points, the velocity of variation, the barycentre, and the deviation from the barycentre are all defined correctly.

Author Contributions: Writing—review & editing, G.P. and M.S. All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are publicly available on the website (http://www.crml.ch, accessed on 1 May 2023).

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

Abbreviations

CoDa	Compositional Data
SRISK	Systemic Risk
IG	Information Geometry

References

1. Kaniadakis, G. Non-linear kinetics underlying generalized statistics. *Phys. A* **2001**, *296*, 405–425. [CrossRef]
2. Kaniadakis, G. H-theorem and generalized entropies within the framework of nonlinear kinetics. *Phys. Lett. A* **2001**, *288*, 283–291. [CrossRef]
3. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Statist. Phys.* **1988**, *52*, 479–487. [CrossRef]
4. Tsallis, C. Introduction to nonextensive statistical mechanics. In *Approaching a Complex World*; Springer: New York, NY, USA, 2009; p. xvii+382.

5. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000; p. x+206. (Translated from the 1993 Japanese original by Daishi Harada)
6. Pistone, G. κ-exponential models from the geometrical viewpoint. *Eur. Phys. J. B Condens. Matter Phys.* **2009**, *71*, 29–37. [CrossRef]
7. Chirco, G.; Pistone, G. Dually affine Information Geometry modeled on a Banach space. *arXiv* **2022**, arXiv:2204.00917.
8. Pawlowsky-Glahn, V.; Egozcue, J.J.; Tolosana-Delgado, R. *Modelling and Analysis of Compositional Data*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2015. [CrossRef]
9. Pawlowsky-Glahn, V.; Egozcue, J.J. Compositional data and their analysis: An introduction. *Geol. Soc. Lond. Spec. Publ.* **2006**, *264*, 1–10. [CrossRef]
10. Egozcue, J.J.; Pawlowsky-Glahn, V. Compositional data: The sample space and its structure. *Test* **2019**, *28*, 599–638. [CrossRef]
11. Coenders, G.; Ferrer-Rosell, B. Compositional data analysis in tourism: Review and future directions. *Tour. Anal.* **2020**, *25*, 153–168. [CrossRef]
12. Fiori, A.M.; Porro, F. A compositional analysis of systemic risk in European financial institutions. *Ann. Financ.* **2023**, *19*, 1–30. [CrossRef]
13. Boonen, T.J.; Guillen, M.; Santolino, M. Forecasting compositional risk allocations. *Insur. Math. Econ.* **2019**, *84*, 79–86. [CrossRef]
14. Grifoll, M.; Ortego, M.; Egozcue, J. Compositional data techniques for the analysis of the container traffic share in a multi-port region. *Eur. Transp. Res. Rev.* **2019**, *11*, 1–15. [CrossRef]
15. Linares-Mustarós, S.; Coenders, G.; Vives-Mestres, M. Financial performance and distress profiles. From classification according to financial ratios to compositional classification. *Adv. Account.* **2018**, *40*, 1–10. [CrossRef]
16. Acharya, V.; Engle, R.; Richardson, M. Capital shortfall: A new approach to ranking and regulating systemic risks. *Am. Econ. Rev.* **2012**, *102*, 59–64. [CrossRef]
17. Acharya, V.V.; Richardson, M.P. (Eds.) *Restoring Financial Stability: How to Repair a Failed System*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2009; Volume 542.
18. Engle, R. Systemic risk 10 years later. *Annu. Rev. Financ. Econ.* **2018**, *10*, 125–152. [CrossRef]
19. Stolbov, M.; Shchepeleva, M. Systemic risk in Europe: Deciphering leading measures, common patterns and real effects. *Ann. Financ.* **2018**, *14*, 49–91. [CrossRef]
20. Engle, R.; Jondeau, E.; Rockinger, M. Systemic risk in Europe. *Rev. Financ.* **2015**, *19*, 145–190. [CrossRef]
21. Aitchison, J. *The Statistical Analysis of Compositional Data*; Monographs on Statistics and Applied Probability; Chapman & Hall: London, UK, 1986; p. xvi+416. [CrossRef]
22. Naudts, J. Generalised exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149. [CrossRef]
23. Pistone, G.; Riccomagno, E.; Wynn, H.P. *Algebraic Statistics: Computational Commutative Algebra in Statistics*; Monographs on Statistics and Applied Probability; Chapman & Hall/CRC: Boca Raton, FL, USA, 2001; Volume 89, p. xvii+160.
24. Montrucchio, L.; Pistone, G. A Class of Non-parametric Deformed Exponential Statistical Models. In *Geometric Structures of Information*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 15–35. [CrossRef]
25. Pistone, G. Information Geometry of the Probability Simplex: A Short Course. *Nonlinear Phenom. Complex Syst.* **2020**, *23*, 221–242. [CrossRef]
26. Kaniadakis, G. Statistical mechanics in the context of special relativity. II. *Phys. Rev. E* **2005**, *72*, 036108. [CrossRef]
27. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **2002**, *66*, 056125. [CrossRef]
28. Montrucchio, L.; Pistone, G. Deformed Exponential Bundle: The Linear Growth Case. In *Lecture Notes in Computer Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 239–246. [CrossRef]
29. Weyl, H. *Space—Time—Matter*; Dover: New York, NY, USA, 1952. (Translation of the 1921 RAUM ZEIT MATERIE)
30. Berger, M. *Geometry I*; Universitext, Springer: Berlin, Germany, 1994; p. xiv+427. (Translated from the 1977 French original by M. Cole and S. Levy, Corrected reprint of the 1987 translation)
31. Pistone, G. Lagrangian Function on the Finite State Space Statistical Bundle. *Entropy* **2018**, *20*, 139. [CrossRef] [PubMed]
32. Chirco, G.; Malagò, L.; Pistone, G. Lagrangian and Hamiltonian dynamics for probabilities on the statistical bundle. *Int. J. Geom. Methods Mod. Phys.* **2022**, *19*, 2250214. [CrossRef]
33. Eguchi, S. Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statist.* **1983**, *11*, 793–803. [CrossRef]