

Article

A Geometric Approach to Average Problems on Multinomial and Negative Multinomial Models

Mingming Li ^{1,2} , Huafei Sun ^{1,2,*} and Didong Li ³ 

¹ School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China; 3120160587@bit.edu.cn

² Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081, China

³ Department of Mathematics, Duke University, Durham, NC 27708, USA; didongli@math.duke.edu

* Correspondence: huafeisun@bit.edu.cn

Received: 12 February 2020; Accepted: 5 March 2020; Published: 8 March 2020



Abstract: This paper is concerned with the formulation and computation of average problems on the multinomial and negative multinomial models. It can be deduced that the multinomial and negative multinomial models admit complementary geometric structures. Firstly, we investigate these geometric structures by providing various useful pre-derived expressions of some fundamental geometric quantities, such as Fisher-Riemannian metrics, α -connections and α -curvatures. Then, we proceed to consider some average methods based on these geometric structures. Specifically, we study the formulation and computation of the midpoint of two points and the Karcher mean of multiple points. In conclusion, we find some parallel results for the average problems on these two complementary models.

Keywords: structure characterization; average problem; geometric midpoints; Karcher mean

1. Introduction

The concept of an average of a set of points within a given geometric structure abounds in various mathematical research. Significant development has been produced since the conception of the Karcher mean [1]. Among the various structure to be studied, the standard simplex presents itself as an interesting framework since it can be directly connected to the parameter spaces of various probability distributions, such as the two to-be-discussed models—the multinomial and negative multinomial models. There are already some recent works involving the statistical modeling of the probability simplex [2]. In the present work, we provide alternative modelings by considering the multinomial and negative multinomial models with classical methods of information geometry.

As a newly developed theory, information geometry has supplied us with various measures of the discrepancy between any two probability distributions. In addition to some standard distance functions, divergence functions are intended for measuring the asymmetric proximity of probability distributions on an appropriate statistical model. Some geometric quantities, such as Riemannian metric and pair of dual connections, can be readily induced from a divergence function by its higher order derivatives [3,4]. In this way, the geometric structures of various parametric statistical models can be studied specifically [5–7], and the investigation about some particular parametric models can also be found in [8,9]. Among these models, we find the multinomial model and negative multinomial model especially interesting, in the sense that they can be seen as a pair of complementary model spaces. The multinomial model is well known as a spherical space of positive constant curvature [10], while the negative multinomial model is found to be a hyperbolic space of negative constant curvature [11].

To be more specific, the motivation of the present paper is twofold. Firstly, we aim at clarifying the complementary geometric structures of the multinomial and negative multinomial models. The main

results about these geometric structures involving geometric quantities, such as Fisher-Riemannian metrics, α -connections and α -curvatures, are collected in Section 3, most of which can be derived in a standard way. In particular, this paper extends the isometric representation results about the multinomial model to those about the negative multinomial model, obtaining new insight into the complementary structures of these two models as illustrated by Table 1. Secondly, the original purpose of the formulation and computation of average problems is approached by utilizing these geometric structures. To this end, we propose a generalized concept of midpoints for two points and a computation scheme of Karcher mean for multiple points. For the midpoints, we generalize the Chernoff points in the literature [12] to some wider parametrized classes. For the Karcher mean, as there are many algorithmic results [13,14] for general manifolds, this paper mainly contributes to addressing some practical issues, such as initial point choice and iteration computation, which yields effective solving methods via the geometric structures within the multinomial model and negative multinomial model. The results about these average methods are presented in Section 4.

2. Preliminaries

For the sake of clarity, we summarize some preliminary knowledge about information divergence functions in this section (more details can be found in [15]).

Given a particular parametric statistical model $\mathcal{M} = \{p_\theta \mid \theta \in \Theta\}$, for our purpose, here we mainly consider invariant divergences that satisfy the property of information monotonicity, as mentioned in [15]. A typical kind of invariant divergence is given in the form of the well-defined f -divergence as

$$D_f(p_\theta \parallel p_\xi) = \int_{x \in \mathcal{X}} p_\theta(x) f\left(\frac{p_\xi(x)}{p_\theta(x)}\right) dx,$$

where f is a convex function satisfying $f(1) = 0$ and $f''(1) = 1$.

A commonly used class of f -divergences is given by the α -divergence $D^{(\alpha)}$, with

$$f(u) = \begin{cases} \frac{4}{1-\alpha^2} \left(1 - u^{\frac{1+\alpha}{2}}\right), & \alpha \neq \pm 1, \\ -\log u, & \alpha = -1, \\ u \log u, & \alpha = 1. \end{cases}$$

Particularly, for $\alpha = -1$, the divergence $D^{(-1)}$ is usually called Kullback–Leibler divergence, which we denote by D_{KL} ; for $\alpha = 0$, $D^{(0)}$ is usually called squared Hellinger distance, which we denote by H^2 ; and for $\alpha = 3$, $D^{(3)}$ is often related to the chi-square statistic.

While our later results are mainly related to the α -divergence, for comparison, we briefly mention another f -divergence called exponential divergence (see [6] §12.3.6), which we denote by \mathcal{E} with $f(u) = \frac{1}{2} \log^2 u$.

For any divergence function D on a statistical manifold, we can construct another divergence function D^* called dual divergence by swapping the arguments as ([16])

$$D^*(p_\theta \parallel p_\xi) = D(p_\xi \parallel p_\theta).$$

The dual divergence of an f -divergence D_f is again an f -divergence $D_f^* = D_{f^*}$, with $f^*(u) = u f(\frac{1}{u})$. Particularly, the dual divergence of the α -divergence $D^{(\alpha)}$ just is $D^{(-\alpha)}$.

To be mentioned, as a divergence is generally not symmetric, a symmetric divergence D_s can be constructed from an asymmetric D by averaging it with its dual:

$$D_s = \frac{1}{2}(D + D^*).$$

A Riemannian metric g can be induced by a divergence D as

$$g \left(\frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \right) = g_{ij}(\theta) = \left. \frac{\partial^2}{\partial \xi_i \partial \xi_j} \right|_{\xi=\theta} D(p_\theta \parallel p_\xi), \tag{1}$$

which is equivalent to the usual Fisher-Riemannian metric in the case when D is an f -divergence.

Furthermore, an affine connection ∇ is induced by the divergence D with connection coefficients

$$g \left(\nabla_{\frac{\partial}{\partial \theta_i}} \frac{\partial}{\partial \theta_j}, \frac{\partial}{\partial \theta_k} \right) = \Gamma_{ijk}(\theta) = - \left. \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \xi_k} \right|_{\xi=\theta} D(p_\theta \parallel p_\xi). \tag{2}$$

Similarly, another affine connection ∇^* can be obtained by replacing D with the dual divergence D^* in the above formula. Thus, with the primal connection ∇ and dual connection ∇^* , the statistical manifold admits a dual structure $(\mathcal{M}, g, \nabla, \nabla^*)$. Furthermore, the structure $(\mathcal{M}, g, \nabla, \nabla^*)$ is called dually flat if both ∇ and ∇^* are flat.

As a well known result [15], the primal connection induced by an f -divergence D_f is the same as the usual α -connection $\nabla^{(\alpha)}$ with $\alpha = 3 + 2 f'''(1)$, while the induced dual connection is $\nabla^{(-\alpha)}$. Particularly, one can check that the primal connection induced by the α -divergence $D^{(\alpha)}$ is exactly the α -connection $\nabla^{(\alpha)}$.

3. Geometric Structure of Multinomial and Negative Multinomial Models

3.1. Basic Information Geometric Structure

In this subsection, we present the parametric formulation of the multinomial and negative multinomial model, respectively. Then some basic results about divergences and geometric structures are derived for both models.

3.1.1. Multinomial Model

Consider the multinomial n -dimensional model \mathcal{M}_N consisting of $(n+1)$ -nominal distributions with probability mass function given by

$$p_\theta(x_0, x_1, \dots, x_n) = \frac{N!}{x_0! x_1! \dots x_n!} \theta_0^{x_0} \theta_1^{x_1} \dots \theta_n^{x_n}, \tag{3}$$

where $\sum_{i=0}^n x_i = N$ and the parametrization is given by $\theta = (\theta_1, \dots, \theta_n)$ with $\theta_0 = 1 - \sum_{i=1}^n \theta_i$.

We can rewrite Equation (3) as

$$p_\theta(x_0, x_1, \dots, x_n) = \frac{N!}{x_0! x_1! \dots x_n!} \exp \left(\sum_{i=1}^n x_i \log \frac{\theta_i}{\theta_0} + N \log \theta_0 \right).$$

By some general knowledge about exponential distribution families [17], we see that the multinomial model \mathcal{M}_N admits the natural parameters

$$\eta_i = \log \frac{\theta_i}{\theta_0}, i = 1, \dots, n, \tag{4}$$

and the potential function

$$\psi(\theta) = -N \log \theta_0 = N \log \left(1 + \sum_{i=1}^n \exp \eta_i \right),$$

from which we also obtain the expectation parameters

$$\partial\psi/\partial\eta_i = N \left(1 + \sum_{i=1}^n \exp \eta_i \right)^{-1} \exp \eta_i = N\theta_i, \quad i = 1, \dots, n. \quad (5)$$

Next, we have the following result by direct calculation.

Proposition 1. *The divergences introduced in Section 2 are obtained for \mathcal{M}_N as follows: the Kullback–Leibler divergence*

$$D_{KL}(p_\theta \| p_\xi) = N \sum_{i=0}^n \theta_i \log \frac{\theta_i}{\xi_i},$$

the squared Hellinger distance

$$H^2(p_\theta, p_\xi) = 4 \left[1 - \left(\sum_{i=0}^n \sqrt{\theta_i \xi_i} \right)^N \right],$$

the α -divergence

$$D^{(\alpha)}(p_\theta \| p_\xi) = \frac{4}{1-\alpha^2} \left[1 - \left(\sum_{i=0}^n \theta_i^{\frac{1-\alpha}{2}} \xi_i^{\frac{1+\alpha}{2}} \right)^N \right]$$

and the exponential divergence

$$\mathcal{E}(p_\theta \| p_\xi) = \frac{N(N-1)}{2} \left(\sum_{i=0}^n \theta_i \log \frac{\theta_i}{\xi_i} \right)^2 + \frac{N}{2} \sum_{i=0}^n \theta_i \log^2 \frac{\theta_i}{\xi_i}.$$

From any one of these expressions, by using Equation (1), we obtain the Fisher-Riemannian metric matrix as

$$g_{ij}(\theta) = \frac{N}{\theta_0} + \frac{N}{\theta_i} \delta_{ij}, \quad (6)$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Then, the inverse matrix of g can also be obtained as

$$g^{ij}(\theta) = \frac{\theta_i}{N} \delta_{ij} - \frac{\theta_i \theta_j}{N}. \quad (7)$$

Furthermore, by direct verification with the metric expression of Equation (6), we have the following well-known result.

Theorem 1 ([18]). *An isometry is established between the multinomial model \mathcal{M}_N and the n -sphere within the non-negative orthant of the Euclidean space \mathbb{R}^{n+1} by the parametric mapping:*

$$(\theta_0, \theta_1, \dots, \theta_n) \mapsto 2\sqrt{N} \left(\theta_0^{1/2}, \theta_1^{1/2}, \dots, \theta_n^{1/2} \right). \quad (8)$$

Consequently, via this isometry, the Fisher-Riemannian geodesic distance between two parameters θ and ξ of \mathcal{M}_N is given by

$$d(\theta, \xi) = 2\sqrt{N} \arccos \left(\sum_{i=0}^n \sqrt{\theta_i \xi_i} \right). \quad (9)$$

Referring to some basic differential geometrical concepts (see [19]), we can derive some further consequent results as follows.

Corollary 1. *The multinomial n -dimensional model \mathcal{M}_N with the Fisher-Riemannian metric is a Riemannian manifold of constant sectional curvature $K = \frac{1}{4N}$ and scalar curvature $S = \frac{n(n-1)}{4N}$. Furthermore, for a unit speed geodesic γ in \mathcal{M}_N , the normal Jacobi fields along γ are precisely linear combinations of the vector fields, which are in the form of $J(t) = \sin(2\sqrt{N}t)E(t)$ or $J(t) = \cos(2\sqrt{N}t)E(t)$, where E is any parallel normal vector field along γ .*

3.1.2. Negative Multinomial Model

Consider the negative multinomial n -dimensional model \mathcal{NM}_M consisting of negative $(n+1)$ -nominal distributions with probability mass function given by

$$p_{\theta}(x_1, \dots, x_n) = \frac{\Gamma(M + x_1 + \dots + x_n)}{\Gamma(M) x_1! \dots x_n!} \theta_0^M \theta_1^{x_1} \dots \theta_n^{x_n}, \quad (10)$$

where $M > 0$, $x_1, \dots, x_n \geq 0$ and the parametrization is given by $\theta = (\theta_1, \dots, \theta_n)$ with $\theta_0 = 1 - \sum_{i=1}^n \theta_i$.

With the rewritten form of Equation (10) as

$$p_{\theta}(x_1, \dots, x_n) = \frac{\Gamma(M + x_1 + \dots + x_n)}{\Gamma(M) x_1! \dots x_n!} \exp\left(\sum_{i=1}^n x_i \log \theta_i + M \log \theta_0\right),$$

we find that the negative multinomial model \mathcal{NM}_M admits the natural parameters

$$\eta_i = \log \theta_i, \quad i = 1, \dots, n, \quad (11)$$

and the potential function

$$\psi(\theta) = -M \log \theta_0 = -M \log\left(1 - \sum_{i=1}^n \exp \eta_i\right),$$

from which we also obtain the expectation parameters

$$\partial\psi/\partial\eta_i = M \left(1 - \sum_{i=1}^n \exp \eta_i\right)^{-1} \exp \eta_i = M \frac{\theta_i}{\theta_0}. \quad (12)$$

Again, we derive the following result by direct calculation.

Proposition 2. *The divergences introduced in Section 2 are obtained for \mathcal{NM}_M as follows: the Kullback–Leibler divergence*

$$D_{KL}(p_{\theta} \parallel p_{\xi}) = M \log \frac{\theta_0}{\xi_0} + M \sum_{i=1}^n \frac{\theta_i}{\theta_0} \log \frac{\theta_i}{\xi_i},$$

the squared Hellinger distance

$$H^2(p_{\theta}, p_{\xi}) = 4 \left[1 - \left(\frac{\sqrt{\theta_0 \xi_0}}{1 - \sum_{i=1}^n \sqrt{\theta_i \xi_i}} \right)^M \right],$$

the α -divergence

$$D^{(\alpha)}(p_{\theta} \parallel p_{\xi}) = \frac{4}{1 - \alpha^2} \left[1 - \left(\frac{\theta_0^{\frac{1-\alpha}{2}} \xi_0^{\frac{1+\alpha}{2}}}{1 - \sum_{i=1}^n \theta_i^{\frac{1-\alpha}{2}} \xi_i^{\frac{1+\alpha}{2}}} \right)^M \right]$$

and the exponential divergence

$$\mathcal{E}(p_\theta \parallel p_\xi) = \frac{M^2}{2} \left(\log \frac{\theta_0}{\xi_0} + \sum_{i=1}^n \frac{\theta_i}{\theta_0} \log \frac{\theta_i}{\xi_i} \right)^2 + \frac{M}{2} \left[\sum_{i=1}^n \frac{\theta_i}{\theta_0} \log^2 \frac{\theta_i}{\xi_i} + \left(\sum_{i=1}^n \frac{\theta_i}{\theta_0} \log \frac{\theta_i}{\xi_i} \right)^2 \right].$$

Next, applying Equation (1) to the divergences in the foregoing proposition, we obtain the Fisher-Riemannian metric matrix as

$$g_{ij} = \frac{M}{\theta_0^2} + \frac{M}{\theta_0 \theta_i} \delta_{ij}, \tag{13}$$

and its inverse matrix as

$$g^{ij} = \frac{\theta_0}{M} (\theta_i \delta_{ij} - \theta_i \theta_j). \tag{14}$$

Furthermore, by direct verification with the metric expression of Equation (13), we have the following result parallel to Theorem 1.

Theorem 2. An isometry is established between the negative multinomial model \mathcal{NM}_M and the n -hyperbola within the nonnegative orthant of the Minkowski space $\mathbb{R}^{1,n}$ by the parametric mapping:

$$(\theta_0, \theta_1, \dots, \theta_n) \mapsto 2\sqrt{M}\theta_0^{-1/2} (1, \theta_1^{1/2}, \dots, \theta_n^{1/2}). \tag{15}$$

Consequently, via this isometry, the Fisher-Riemannian geodesic distance between two parameters θ and ξ of \mathcal{NM}_M is given by

$$d(\theta, \xi) = 2\sqrt{M} \operatorname{arcosh} \left(\frac{1 - \sum_{i=1}^n \sqrt{\theta_i \xi_i}}{\sqrt{\theta_0 \xi_0}} \right). \tag{16}$$

Again, some further consequent results are obtained as follows.

Corollary 2. The negative multinomial n -dimensional model \mathcal{NM}_M with the Fisher-Riemannian metric is a Riemannian manifold of constant sectional curvature $K = -\frac{1}{4M}$ and scalar curvature $S = -\frac{n(n-1)}{4M}$. Furthermore, for a unit speed geodesic γ in \mathcal{NM}_M , the normal Jacobi fields along γ are precisely linear combinations of the vector fields, which are in the form of $J(t) = \sinh(2\sqrt{M}t)E(t)$ or $J(t) = \cosh(2\sqrt{M}t)E(t)$, where E is any parallel normal vector field along γ .

3.2. Dual Structures

In this section, we derive the α -connection coefficients and curvatures of the multinomial and negative multinomial model. While some basic results are equivalent to those in [11], our calculation is performed directly with the original parameters in order to give a clear presentation of the results.

The α -connection coefficients can be obtained by applying the calculation of Equation (2) to the α -divergence $D^{(\alpha)}$, as mentioned in Section 2. However, an easier derivation of the α -connection $\nabla^{(\alpha)}$ is given in the form of a linear combination of the mixture connection $\nabla^{(m)} = \nabla^{(-1)}$ and the exponential connection $\nabla^{(e)} = \nabla^{(1)}$. With the mixture connection coefficients obtained from the Kullback–Leibler divergence as

$$\Gamma_{ijk}^{(m)}(\theta) = - \left. \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \xi_k} \right|_{\xi=\theta} D_{KL}(p_\theta \parallel p_\xi), \tag{17}$$

and the exponential connection coefficients obtained from the dual Kullback–Leibler divergence as

$$\Gamma_{ijk}^{(e)}(\theta) = - \left. \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \xi_k} \right|_{\xi=\theta} D_{KL}^*(p_\theta \parallel p_\xi), \tag{18}$$

the α -connection coefficients are given by

$$\Gamma_{ijk}^{(\alpha)}(\theta) = \frac{1 + \alpha}{2} \Gamma_{ijk}^{(e)}(\theta) + \frac{1 - \alpha}{2} \Gamma_{ijk}^{(m)}(\theta). \tag{19}$$

Next, the α -curvature tensor $R^{(\alpha)}$ is defined by

$$R^{(\alpha)}(X, Y)Z = \nabla_X^{(\alpha)} \nabla_Y^{(\alpha)} Z - \nabla_Y^{(\alpha)} \nabla_X^{(\alpha)} Z - \nabla_{[X, Y]}^{(\alpha)} Z,$$

where $[X, Y]$ denotes the Lie bracket of X and Y . By the duality condition (see [20])

$$Xg(Y, Z) = g(\nabla_X^{(\alpha)} Y, Z) + g(Y, \nabla_X^{(-\alpha)} Z),$$

one can check that the following identity holds:

$$g\left(R^{(\alpha)}(\partial_i, \partial_j)\partial_j, \partial_i\right) = \partial_i \Gamma_{jji}^{(\alpha)} - \partial_j \Gamma_{iji}^{(\alpha)} + \sum_{k,l} g^{kl} \left(\Gamma_{ijk}^{(\alpha)} \Gamma_{jil}^{(-\alpha)} - \Gamma_{jjk}^{(\alpha)} \Gamma_{iil}^{(-\alpha)}\right), \tag{20}$$

where $\frac{\partial}{\partial \theta_i}$ is shortened as ∂_i .

At last, the α -sectional curvature spanned by two tangent vectors ∂_i and ∂_j ($i \neq j$) is determined by

$$K^{(\alpha)}(\partial_i, \partial_j) = \frac{g\left(R^{(\alpha)}(\partial_i, \partial_j)\partial_j, \partial_i\right)}{g_{ii}g_{jj} - g_{ij}^2}. \tag{21}$$

3.2.1. Multinomial Model

For the multinomial model \mathcal{M}_N , by applying Equations (17)–(19) with the Kullback–Leibler divergence in Proposition 1, we have the mixture connection coefficients

$$\Gamma_{ijk}^{(m)}(\theta) = 0,$$

the exponential connection coefficients

$$\Gamma_{ijk}^{(e)}(\theta) = \frac{N}{\theta_0^2} - \frac{N}{\theta_i^2} \delta_{ijk}$$

and the α -connection coefficients

$$\Gamma_{ijk}^{(\alpha)}(\theta) = \frac{1 + \alpha}{2} \Gamma_{ijk}^{(e)}(\theta).$$

Furthermore, by Equations (6), (7) and (20), we obtain

$$g\left(R^{(\alpha)}(\partial_i, \partial_j)\partial_j, \partial_i\right) = \frac{(1 - \alpha^2)N}{4} \left(\frac{1}{\theta_0\theta_i} + \frac{1}{\theta_0\theta_j} + \frac{1}{\theta_i\theta_j}\right),$$

$$g_{ii}g_{jj} - g_{ij}^2 = N^2 \left(\frac{1}{\theta_0\theta_i} + \frac{1}{\theta_0\theta_j} + \frac{1}{\theta_i\theta_j}\right).$$

Thus, via Equation (21), we recover the following result.

Theorem 3 ([11]). *The multinomial model \mathcal{M}_N admits constant α -sectional curvature*

$$K^{(\alpha)} = \frac{1 - \alpha^2}{4N}. \tag{22}$$

3.2.2. Negative Multinomial Model

For the negative multinomial model \mathcal{NM}_M , by applying Equations(17)–(19) with the Kullback–Leibler divergence in Proposition 2, we have the mixture connection coefficients

$$\Gamma_{ijk}^{(m)}(\theta) = \frac{M}{\theta_0^2} \left(\frac{2}{\theta_0} + \frac{\delta_{ik} + \delta_{jk}}{\theta_k} \right) = (g_{ik} + g_{jk})/\theta_0,$$

the exponential connection coefficients

$$\Gamma_{ijk}^{(e)}(\theta) = -\frac{M}{\theta_0\theta_i^2}\delta_{ijk} - \frac{M}{\theta_0^2\theta_i}\delta_{ij} = -g_{ik}\delta_{ij}/\theta_i$$

and the α -connection coefficients

$$\Gamma_{ijk}^{(\alpha)}(\theta) = \frac{1 + \alpha}{2}\Gamma_{ijk}^{(e)}(\theta) + \frac{1 - \alpha}{2}\Gamma_{ijk}^{(m)}(\theta).$$

Furthermore, by Equations (13), (14) and (20), we have

$$g \left(R^{(\alpha)}(\partial_i, \partial_j)\partial_j, \partial_i \right) = -\frac{(1 - \alpha^2)M}{4\theta_0^2} \left(\frac{1}{\theta_0\theta_i} + \frac{1}{\theta_0\theta_j} + \frac{1}{\theta_i\theta_j} \right),$$

$$g_{ii}g_{jj} - g_{ij}^2 = \frac{M^2}{\theta_0^2} \left(\frac{1}{\theta_0\theta_i} + \frac{1}{\theta_0\theta_j} + \frac{1}{\theta_i\theta_j} \right).$$

Again, via Equation (21), we recover another parallel result.

Theorem 4 ([11]). *The negative multinomial model \mathcal{NM}_M admits constant α -sectional curvature*

$$K^{(\alpha)} = -\frac{1 - \alpha^2}{4M}. \tag{23}$$

For clarity, we summarize these results about the complementary geometric structures of the multinomial and negative multinomial models in Table 1.

Table 1. Complementary geometric structures.

Model	Isometric Space	Sectional Curvature	α -Sectional Curvature
\mathcal{M}_N	n -sphere within nonnegative orthant	$\frac{1}{4N}$	$\frac{1-\alpha^2}{4N}$
\mathcal{NM}_M	n -hyperbola within nonnegative orthant	$-\frac{1}{4M}$	$-\frac{1-\alpha^2}{4M}$

4. Geometric Average Methods on Multinomial and Negative Multinomial Models

In this section, we present some average methods induced by the geometry of the multinomial and negative multinomial models.

Firstly, we consider the particular case when the to-be-averaged set consists of only two points. In this case, the problem is to find a method of computing a midpoint in some geometric sense. Next, via some techniques related to the Karcher mean, we consider the general case with a set of multiple points.

4.1. Midpoints of Two Points

In this subsection, again within the multinomial and negative multinomial model, we study a particular class of midpoints named Chernoff points. The original Chernoff point, which is motivated by the application of computing the best error exponent for the Bayesian hypothesis

testing problem, is determined as the intersection point of an exponential geodesic and a mixture bisector [21]. Furthermore, there are three other generalized Chernoff points proposed by [12].

To present a further generalization, here we formulate the concepts of α -geodesic and α -bisector determined by two probability distributions $p_{\theta'}$ and $p_{\theta''}$ of a parametric statistical model \mathcal{M} .

Definition 1. The α -geodesic is determined by the geodesic equation of the α -connection $\nabla^{(\alpha)}$ as

$$G^{(\alpha)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta(t)} \in \mathcal{M} \mid \theta(0) = \theta', \theta(1) = \theta'', \nabla_{\dot{\theta}(t)}^{(\alpha)} \dot{\theta}(t) = 0, t \in [0, 1] \right\},$$

where $\dot{\theta}(t)$ denotes the velocity vector of a curve $\theta(t)$.

Particularly, since an exponential family model is ± 1 -flat, as can be directly seen from Equations (22) and (23) for our case, the exponential geodesic for $\alpha = 1$ can be determined by the linear interpolation of the natural parameters, while the mixture geodesic for $\alpha = -1$ can be determined by the linear interpolation of the expectation parameters.

Definition 2. The α -bisector is determined by the equi-divergence identity of the α -divergence $D^{(\alpha)}$ as

$$\text{Bi}^{(\alpha)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta} \in \mathcal{M} \mid D^{(\alpha)}(p_{\theta} \parallel p_{\theta'}) = D^{(\alpha)}(p_{\theta} \parallel p_{\theta''}) \right\}. \quad (24)$$

Particularly, the exponential bisector for $\alpha = 1$ is determined in terms of the dual Kullback–Leibler divergence, while the mixture bisector for $\alpha = -1$ is determined in terms of the Kullback–Leibler divergence.

Then, we can generalize the notion of Chernoff points suggested by [12] as follows.

Definition 3. Two types of generalized Chernoff points are given by the intersection points with parameter α :

$$CP_{\text{I}}^{(\alpha)}(p_{\theta'}, p_{\theta''}) = G^{(\alpha)}(p_{\theta'}, p_{\theta''}) \cap \text{Bi}^{(-\alpha)}(p_{\theta'}, p_{\theta''}),$$

$$CP_{\text{II}}^{(\alpha)}(p_{\theta'}, p_{\theta''}) = G^{(\alpha)}(p_{\theta'}, p_{\theta''}) \cap \text{Bi}^{(\alpha)}(p_{\theta'}, p_{\theta''}).$$

Thus, the Chernoff points already proposed in previous works can be recovered by setting $\alpha = \pm 1$ in Definition 3.

The existence of these intersection points is assured by the intermediate value property of the determining Equation (24), since replacing θ by θ' and θ'' , respectively, we get two opposite inequalities due to the non-negativity of the α -divergence.

While the uniqueness can be proven for an exponential family model if $\alpha = \pm 1$ (see [12]), we conjecture it still holds for general cases but this is not pursued in the present paper.

To elucidate what we have mentioned earlier about the application of the original Chernoff point ($CP_{\text{I}}^{(1)}$ in our notation) to the binary Bayesian hypothesis testing problem, we present here the upper bound of the probability of error of the Bayesian decision suggested by [21] as

$$\exp(-D_{\text{KL}}(p_{\theta^*} \parallel p_{\theta'})) = \exp(-D_{\text{KL}}(p_{\theta^*} \parallel p_{\theta''})), \quad (25)$$

where p_{θ^*} denotes the Chernoff point $CP_{\text{I}}^{(1)}(p_{\theta'}, p_{\theta''})$.

The overlapping case of the two classes of generalized Chernoff points is given by $\alpha = 0$. For the multinomial and negative multinomial model, by comparing the 0-divergence, i.e., squared Hellinger distance, with the Fisher-Riemannian geodesic distance presented in Section 3.1, we find that the generalized Chernoff point $CP_{\text{I}}^{(0)}(p_{\theta'}, p_{\theta''}) = CP_{\text{II}}^{(0)}(p_{\theta'}, p_{\theta''})$ is exactly the unique Fisher-Riemannian geodesic midpoint between $p_{\theta'}$ and $p_{\theta''}$ within both models.

Next, we summarize some specific results about Chernoff points for both models as follows.

4.1.1. Multinomial Model

For the multinomial model \mathcal{M}_N , although the geodesic equation of the α -geodesic can be explicitly given in general cases, there are simple closed-form geodesic expressions at least for $\alpha = \pm 1$.

Proposition 3. *The exponential and mixture geodesics connecting two probability distributions $p_{\theta'}$ and $p_{\theta''}$ of the multinomial model \mathcal{M}_N are given by*

$$G^{(1)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta(t)} \in \mathcal{M}_N \mid \theta_i(t) = \frac{(\theta'_i)^{1-t} (\theta''_i)^t}{\sum_{i=0}^n (\theta'_i)^{1-t} (\theta''_i)^t}, i = 0, \dots, n, t \in [0, 1] \right\},$$

$$G^{(-1)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta(t)} \in \mathcal{M}_N \mid \theta_i(t) = (1-t)\theta'_i + t\theta''_i, i = 0, \dots, n, t \in [0, 1] \right\}.$$

Proof. As already mentioned, the exponential and mixture geodesics can be easily obtained by the linear interpolation of the natural and expectation parameters by Equations (4) and (5), respectively. \square

By using the expressions of α -divergences presented in Proposition 1, the α -bisectors are directly obtained as follows.

Proposition 4. *The α -bisectors between $p_{\theta'}$ and $p_{\theta''}$ within the multinomial model \mathcal{M}_N are given by the following equations:*

$$\text{Bi}^{(\alpha)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta} \in \mathcal{M}_N \mid \sum_{i=0}^n \theta_i^{\frac{1-\alpha}{2}} \left[(\theta'_i)^{\frac{1+\alpha}{2}} - (\theta''_i)^{\frac{1+\alpha}{2}} \right] = 0 \right\}, \alpha \neq \pm 1,$$

$$\text{Bi}^{(1)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta} \in \mathcal{M}_N \mid \sum_{i=0}^n (\theta'_i - \theta''_i) \log \theta_i = \sum_{i=0}^n (\theta'_i \log \theta'_i - \theta''_i \log \theta''_i) \right\},$$

$$\text{Bi}^{(-1)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta} \in \mathcal{M}_N \mid \sum_{i=0}^n \theta_i \log \frac{\theta'_i}{\theta''_i} = 0 \right\}.$$

Combining the previous two propositions, we have the following result about the determining equations for the four particular Chernoff points with $\alpha = \pm 1$.

Theorem 5. *The determining equations for the Chernoff points with $\alpha = \pm 1$ of $p_{\theta'}$ and $p_{\theta''}$ within the multinomial model \mathcal{M}_N are expressed in terms of the argument $t \in [0, 1]$ of the corresponding geodesics as follows:*

1. For $CP_I^{(1)}(p_{\theta'}, p_{\theta''})$,

$$\sum_{i=0}^n (\theta'_i)^{1-t} (\theta''_i)^t \log \frac{\theta'_i}{\theta''_i} = 0;$$

2. For $CP_I^{(-1)}(p_{\theta'}, p_{\theta''})$,

$$\sum_{i=0}^n (\theta'_i - \theta''_i) \log [(1-t)\theta'_i + t\theta''_i] = \sum_{i=0}^n (\theta'_i \log \theta'_i - \theta''_i \log \theta''_i);$$

3. For $CP_{II}^{(1)}(p_{\theta'}, p_{\theta''})$,

$$t = \frac{D_{KL}(p_{\theta''} \parallel p_{\theta'})}{D_{KL}(p_{\theta'} \parallel p_{\theta''}) + D_{KL}(p_{\theta''} \parallel p_{\theta'})};$$

4. For $CP_{II}^{(-1)}(p_{\theta'}, p_{\theta''})$,

$$t = \frac{D_{KL}(p_{\theta'} \parallel p_{\theta''})}{D_{KL}(p_{\theta'} \parallel p_{\theta''}) + D_{KL}(p_{\theta''} \parallel p_{\theta'})}.$$

With the Kullback–Leibler divergence D_{KL} given by Proposition 1, the two points of second type $CP_{II}^{(1)}$ and $CP_{II}^{(-1)}$ are already in an explicit form. Whereas, the two points of first type $CP_I^{(1)}$ and $CP_I^{(-1)}$ are to be solved by using some numerical methods such as simple bisection as suggested in [12].

For the Fisher-Riemannian geodesic midpoint, we have the following result.

Theorem 6. *The Fisher-Riemannian geodesic midpoint p_{θ^*} between $p_{\theta'}$ and $p_{\theta''}$ in the multinomial model \mathcal{M}_N is determined by*

$$\theta_i^* = \frac{(\sqrt{\theta'_i} + \sqrt{\theta''_i})^2}{2 + 2 \sum_{i=0}^n \sqrt{\theta'_i \theta''_i}}, i = 0, \dots, n.$$

Proof. Let F be the isometry given by Equation (8). Denote the linear midpoint of the two image points $\frac{F(\theta') + F(\theta'')}{2}$ by $x^* \in \mathbb{R}^{n+1}$. Then we normalize x^* to the n -sphere as $u^* = 2\sqrt{N}x^* / \|x^*\|_e$, where the Euclidean norm $\|x\|_e = (\sum_{i=0}^n x_i^2)^{1/2}$ is used. At last, the required midpoint is obtained as the inverse image point $\theta^* = F^{-1}(u^*)$. □

To illustrate these notions for the multinomial model \mathcal{M}_N , we present a numerical example as follows. The two parameters θ' and θ'' are taken as the empirical probability vectors of the first and second 100 decimal digits of π , respectively. The parameters of these two points and the resulting Chernoff points are summarized in Table 2.

Table 2. Chernoff points of multinomial model (%).

	0	1	2	3	4	5	6	7	8	9
θ'	8.00	8.00	12.00	11.00	10.00	8.00	9.00	8.00	12.00	14.00
θ''	11.00	12.00	12.00	8.00	12.00	12.00	7.00	4.00	13.00	9.00
$CP_I^{(1)}$	9.49	9.91	12.17	9.53	11.10	9.91	8.06	5.76	12.66	11.41
$CP_I^{(-1)}$	9.52	10.02	12.00	9.48	11.01	10.02	7.99	5.98	12.51	11.47
$CP_{II}^{(1)}$	9.48	9.89	12.17	9.55	11.08	9.89	8.07	5.78	12.65	11.44
$CP_{II}^{(-1)}$	9.54	10.05	12.00	9.46	11.02	10.05	7.98	5.95	12.51	11.44
$CP^{(0)}$	9.51	9.97	12.08	9.51	11.05	9.97	8.02	5.87	12.58	11.44

As we can see, the pair of points $CP_I^{(1)}$ and $CP_{II}^{(1)}$ admits a certain similarity between them as both of them lie on the same exponential geodesic, and so does the pair of points $CP_I^{(-1)}$ and $CP_{II}^{(-1)}$ as both of them lie on the same mixture geodesic. Whereas, the Fisher-Riemannian geodesic midpoint $CP^{(0)}$ can be considered as a medium version among these Chernoff points.

To be mentioned particularly, the upper bound of the probability of error given by Equation (25) is obtained via $CP_I^{(1)}$ as being equal to 0.9862^N . Thus, we can choose sufficiently large N so that the probability of error is less than some threshold value.

4.1.2. Negative Multinomial Model

For the negative multinomial model \mathcal{NM}_M , again we can give the geodesic expressions for the exponential and mixture cases.

Proposition 5. *The exponential and mixture geodesics connecting two probability distributions $p_{\theta'}$ and $p_{\theta''}$ of the negative multinomial model \mathcal{NM}_M are given by*

$$G^{(1)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta(t)} \in \mathcal{NM}_M \mid \theta_i(t) = (\theta'_i)^{1-t} (\theta''_i)^t, i = 1, \dots, n, t \in [0, 1] \right\},$$

$$G^{(-1)}(p_{\theta'}, p_{\theta''}) = \left\{ p_{\theta(t)} \in \mathcal{NM}_M \mid \theta_i(t) = \frac{(1-t)\frac{\theta'_i}{\theta'_0} + t\frac{\theta''_i}{\theta''_0}}{\frac{1-t}{\theta'_0} + \frac{t}{\theta''_0}}, i = 1, \dots, n, t \in [0, 1] \right\}.$$

Proof. By using Equations (11) and (12), the exponential and mixture geodesics can be easily obtained by the linear interpolation of the natural and expectation parameters, respectively. \square

By using the expressions of α -divergences presented in Proposition 2, the α -bisectors are directly obtained as follows.

Proposition 6. The α -bisectors between $p_{\theta'}$ and $p_{\theta''}$ within the negative multinomial model \mathcal{NM}_M are given by the following equations:

$$\begin{aligned} \text{Bi}^{(\alpha)}(p_{\theta'}, p_{\theta''}) &= \left\{ p_{\theta} \in \mathcal{M}_N \mid \sum_{i=1}^n \theta_i^{\frac{1-\alpha}{2}} \left[\left(\frac{\theta'_i}{\theta'_0} \right)^{\frac{1+\alpha}{2}} - \left(\frac{\theta''_i}{\theta''_0} \right)^{\frac{1+\alpha}{2}} \right] = (\theta'_0)^{-\frac{1+\alpha}{2}} - (\theta''_0)^{-\frac{1+\alpha}{2}} \right\}, \\ \text{Bi}^{(1)}(p_{\theta'}, p_{\theta''}) &= \left\{ p_{\theta} \in \mathcal{NM}_M \mid \sum_{i=1}^n \left(\frac{\theta'_i}{\theta'_0} - \frac{\theta''_i}{\theta''_0} \right) \log \theta_i = \sum_{i=0}^n \left(\frac{\theta'_i}{\theta'_0} \log \theta'_i - \frac{\theta''_i}{\theta''_0} \log \theta''_i \right) \right\}, \\ \text{Bi}^{(-1)}(p_{\theta'}, p_{\theta''}) &= \left\{ p_{\theta} \in \mathcal{NM}_M \mid \sum_{i=0}^n \theta_i \log \frac{\theta'_i}{\theta''_i} = 0 \right\}. \end{aligned}$$

Combining the previous two propositions, we have the following result about the determining equations for the four particular Chernoff points with $\alpha = \pm 1$.

Theorem 7. The determining equations for the Chernoff points with $\alpha = \pm 1$ of $p_{\theta'}$ and $p_{\theta''}$ within the negative multinomial model \mathcal{NM}_M are expressed in terms of the argument $t \in [0, 1]$ of the corresponding geodesics as follows:

1. For $CP_I^{(1)}(p_{\theta'}, p_{\theta''})$,

$$\log \frac{\theta'_0}{\theta''_0} + \sum_{i=1}^n (\theta'_i)^{1-t} (\theta''_i)^t \left(\log \frac{\theta'_i}{\theta''_i} - \log \frac{\theta'_0}{\theta''_0} \right) = 0;$$

2. For $CP_I^{(-1)}(p_{\theta'}, p_{\theta''})$,

$$\sum_{i=1}^n \left(\frac{\theta'_i}{\theta'_0} - \frac{\theta''_i}{\theta''_0} \right) \log \frac{(1-t)\frac{\theta'_i}{\theta'_0} + t\frac{\theta''_i}{\theta''_0}}{\frac{1-t}{\theta'_0} + \frac{t}{\theta''_0}} = \sum_{i=0}^n \left(\frac{\theta'_i}{\theta'_0} \log \theta'_i - \frac{\theta''_i}{\theta''_0} \log \theta''_i \right);$$

3. For $CP_{II}^{(1)}(p_{\theta'}, p_{\theta''})$,

$$t = \frac{D_{KL}(p_{\theta''} \parallel p_{\theta'})}{D_{KL}(p_{\theta'} \parallel p_{\theta''}) + D_{KL}(p_{\theta''} \parallel p_{\theta'})};$$

4. For $CP_{II}^{(-1)}(p_{\theta'}, p_{\theta''})$,

$$t = \frac{D_{KL}(p_{\theta'} \parallel p_{\theta''})}{D_{KL}(p_{\theta'} \parallel p_{\theta''}) + D_{KL}(p_{\theta''} \parallel p_{\theta'})}.$$

As can be seen, the two points of second type $CP_{II}^{(1)}$ and $CP_{II}^{(-1)}$ are in the same explicit form as before, except that the Kullback–Leibler divergence D_{KL} is given by Proposition 2. And again, the two points of first type $CP_I^{(1)}$ and $CP_I^{(-1)}$ can be solved by using numerical methods.

For the Fisher-Riemannian geodesic midpoint, we have the following result being complementary to the previous one.

Theorem 8. The Fisher-Riemannian geodesic midpoint p_{θ^*} between $p_{\theta'}$ and $p_{\theta''}$ in the negative multinomial model $\mathcal{N}\mathcal{M}_M$ is determined by

$$\theta_i^* = \left[\frac{(\theta'_0)^{-1/2} (\theta'_i)^{1/2} + (\theta''_0)^{-1/2} (\theta''_i)^{1/2}}{(\theta'_0)^{-1/2} + (\theta''_0)^{-1/2}} \right]^2, i = 1, \dots, n.$$

Proof. Let F be the isometry given by Equation (15). Denote the linear midpoint of the two image points $\frac{F(\theta') + F(\theta'')}{2}$ by $x^* \in \mathbb{R}^{1,n}$. Then we normalize x^* to the n -hyperbola as $u^* = 2\sqrt{M} x^* / \|x^*\|_m$, where the Minkowski norm $\|x\|_m = (x_0^2 - \sum_{i=1}^n x_i^2)^{1/2}$ is used. At last, the required midpoint is obtained as the inverse image point $\theta^* = F^{-1}(u^*)$. \square

A numerical illustration for the negative multinomial model $\mathcal{N}\mathcal{M}_M$ is presented as follows. The two parameters θ' and θ'' are taken as the empirical probability vectors of the decimal digits of π within the first and second 10 appearances of "0", respectively. The parameters of these two points and the resulting Chernoff points are summarized in Table 3.

Table 3. Chernoff points of negative multinomial model (%).

	0	1	2	3	4	5	6	7	8	9
θ'	8.62	8.62	12.93	11.21	9.48	7.76	8.62	6.90	13.79	12.07
θ''	10.99	12.09	10.99	6.59	14.29	12.09	6.59	4.40	12.09	9.89
$CP_I^{(1)}$	10.99	10.11	11.98	8.73	11.50	9.56	7.60	5.58	12.96	10.99
$CP_I^{(-1)}$	9.73	10.25	12.02	9.04	11.74	9.79	7.67	5.72	12.99	11.05
$CP_{II}^{(1)}$	10.91	10.00	12.04	8.87	11.36	9.43	7.66	5.66	13.01	11.06
$CP_{II}^{(-1)}$	9.80	10.35	11.96	8.90	11.88	9.92	7.61	5.65	12.94	10.98
$CP^{(0)}$	10.36	10.18	12.00	8.89	11.62	9.67	7.63	5.65	12.98	11.02

Again, the pair of points $CP_I^{(1)}$ and $CP_{II}^{(1)}$ admits a certain similarity between them, and so does the pair of points $CP_I^{(-1)}$ and $CP_{II}^{(-1)}$. Whereas, the Fisher-Riemannian geodesic midpoint $CP^{(0)}$ serves as a medium version among these Chernoff points.

The upper bound of the probability of error given by Equation (25) is obtained via $CP_I^{(1)}$ as being equal to 0.8789^M . Thus, sufficiently large M need to be chosen so that the probability of error is less than some threshold value.

4.2. Karcher Means of Multiple Points

A natural generalization of the Fisher-Riemannian geodesic midpoint between two points is given by the Karcher mean among multiple points.

Let \mathcal{M} be a metric space and S be a set of points on \mathcal{M} . Define a criterion function $f : \mathcal{M} \rightarrow \mathbb{R}$ by

$$f(x) := \frac{1}{2|S|} \sum_{p \in S} d(x, p)^2,$$

where $d(\cdot, \cdot)$ is the distance function and $|S|$ is the number of points of S . If the minimizer of the function f exists and is unique, then it is called the *Karcher mean* of S on \mathcal{M} . If d is a Riemannian metric on \mathcal{M} , then the negative gradient vector field with respect to f is found to be the usual average of the corresponding points in the tangent space ([1]):

$$-\nabla f(x) = \frac{1}{|S|} \sum_{p \in S} \exp_x^{-1}(p), \tag{26}$$

where \exp_x^{-1} is the inverse of the Riemannian exponential map at x . In view of this, the Karcher mean can be alternatively understood as a point at which the above vector field vanishes.

The Karcher mean may be not unique unless all points are located on a geodesically convex region. For example, there are infinitely many geodesic midpoints between two antipodal points on a sphere. However, for model spaces, such as open half-sphere and hyperbolic space, there are existing results to assure the existence and uniqueness of Karcher mean ([22]). Thus, by virtue of Theorem 1 and Theorem 2, we conclude that the concept of a Karcher mean is well-defined on the multinomial and negative multinomial models.

Now, we focus ourselves on the computation of the Karcher mean on these two models. The Karcher mean of two points admits a closed-form expression as the Fisher-Riemannian geodesic midpoint presented previously, but for multiple points, we can only expect to obtain a numerical solution of the Karcher mean.

By virtue of Equation (26), there is a Riemannian gradient iteration algorithm with a locally superlinear convergence in general ([23]):

$$x_{i+1} = \exp_{x_i} \left(\frac{1}{|S|} \sum_{p \in S} \exp_{x_i}^{-1}(p) \right).$$

However, this general algorithm is difficult to apply in practice unless proper representations of the models are derived. In our case, as we have prepared enough geometric representation results within the multinomial and negative multinomial models in Section 3, we still have to address two practical issues: the choice of initial points and the computation of Riemannian exponential map \exp_{x_i} and its inverse $\exp_{x_i}^{-1}$.

4.2.1. Initial Points

Let S be a set of parameters to be averaged in either the model \mathcal{M}_N or \mathcal{NM}_M , we present a heuristic approach motivated by the proof of Theorem 1 and Theorem 2 to provide an initial point choice. The main procedure is presented as follows (here $N = M = 1$ is assumed as basic ideas are unchanged up to scale):

1. Set the average of isometry images $x^* := \frac{1}{|S|} \sum_{\theta \in S} F(\theta)$;
2. Set the normalized vector $u^* := 2x^* / \|x^*\|$; the parameter of the initial point is given by $\theta^* := F^{-1}(u^*)$.

For the model \mathcal{M}_N , the isometry F is given by Equation (8), and the norm $\|\cdot\|$ is the Euclidean norm $\|\cdot\|_e$ in the proof of Theorem 1. For the model \mathcal{NM}_M , the isometry F is given by Equation (15), and the norm $\|\cdot\|$ is the Minkowski norm $\|\cdot\|_m$ in the proof of Theorem 2.

4.2.2. Computation of Riemannian Exponential and its Inverse

Within each of the model \mathcal{M}_N and \mathcal{NM}_M (again $N = M = 1$ is assumed), the Riemannian exponential and its inverse map can be computed in an easy-to-manipulate way. Except for the isometry F and the norm $\|\cdot\|$ being given as before, we also need to set $\langle \cdot, \cdot \rangle$ by $\langle x, y \rangle = \sum_{i=0}^n x_i y_i$ as the Euclidean inner product for the model \mathcal{M}_N and by $\langle x, y \rangle = x_0 y_0 - \sum_{i=1}^n x_i y_i$ as the Minkowski inner product for the model \mathcal{NM}_M .

To compute $\exp_{\theta}^{-1}(\xi)$, we have the following steps:

1. Set a tangent vector by orthogonal projection

$$x := F(\xi) - F(\theta) \langle F(\xi), F(\theta) \rangle / 4;$$

2. Scale the vector by geodesic distance

$$u := d(\theta, \xi) x / \|x\|,$$

where the geodesic distance $d(\cdot, \cdot)$ is given by Equation (9) for \mathcal{M}_N and by Equation (16) for \mathcal{NM}_M .

Thus, we use the resulting vector u to represent $\exp_{\theta}^{-1}(\xi)$.

To compute $\exp_{\theta}(u)$, we need to:

1. Set the angle $\alpha := \|u\|/2$;
2. For \mathcal{M}_N , express the corresponding point on the sphere

$$x := F(\theta) \cos \alpha + 2 \|u\|^{-1} u \sin \alpha,$$

for \mathcal{NM}_M , express the corresponding point on the hyperbola

$$x := F(\theta) \cosh \alpha + 2 \|u\|^{-1} u \sinh \alpha;$$

3. Set the parameter by isometry $\xi := F^{-1}(x)$.

Thus, the resulting parameter ξ is obtained as $\exp_{\theta}(u)$.

4.2.3. Numerical Example

Now, we test the above algorithm for solving Karcher mean by a numerical example. The data set S is chosen as containing 10 empirical probability vectors from the first to the tenth 100 decimal digits of π .

To illustrate the goodness of each iteration, we present the norm of the negative gradient vector field at each iteration point via Equation (26), as shown in Table 4. Within each model, we present in column (a) the iteration results starting with the initial points chosen as the aforementioned way, while in column (b), the iteration results with initial points chosen as the usual Euclidean means are provided for comparison.

Table 4. Iteration for the Karcher mean illustrated by $\|-\nabla f\|$.

Iteration	\mathcal{M}_N		\mathcal{NM}_M	
	(a)	(b)	(a)	(b)
0	8.21×10^{-5}	1.07×10^{-2}	3.05×10^{-3}	1.85×10^{-1}
1	4.62×10^{-7}	6.28×10^{-5}	1.74×10^{-4}	1.20×10^{-2}
2	2.68×10^{-9}	3.90×10^{-7}	1.01×10^{-5}	7.88×10^{-4}

As we can see, all of the four iterations shown here converge rapidly within the first two steps, and our aforementioned choice for initial points is apparently better than the usual choice of Euclidean means. In conclusion, this example, to some extent, shows the effectiveness of our computation scheme for the Karcher mean within the multinomial and negative multinomial models.

5. Conclusions

In this paper, we have studied various information geometric properties based on divergence functions for the multinomial and negative multinomial models. Some pre-derived expressions of fundamental geometric quantities, such as Fisher-Riemannian metric, isometric representation and α -curvature, have made it clear that these two models can be put together into a complementary view. With the aid of these geometric structures, we investigate the average problems on these two models. We have proposed the conception of generalized Chernoff points as midpoints of two points and presented some determining equations for them. Then we provided an effective computation scheme for the Karcher mean of multiple points on the multinomial and negative multinomial models.

Author Contributions: Conceptualization, M.L. and H.S.; funding acquisition, H.S.; supervision, H.S.; validation, M.L., H.S. and D.L.; writing—original draft, M.L.; writing—review and editing, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 61179031).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Karcher, H. Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.* **1977**, *30*, 509–541. [[CrossRef](#)]
2. Nielsen, F.; Sun, K. Clustering in Hilbert’s projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices. *Geom. Struct. Inf.* **2019**. [[CrossRef](#)]
3. Eguchi, S. Second order efficiency of minimum contrast estimations in a curved exponential family. *Ann. Stat.* **1983**, *11*, 793–803. [[CrossRef](#)]
4. Eguchi, S. A characterization of second order efficiency in a curved exponential family. *Ann. Inst. Stat. Math.* **1984**, *36*, 199–206. [[CrossRef](#)]
5. Arwini, K.A.; Dodson, C.T. *Information Geometry: Near Randomness and Near Independence*; Springer: Berlin, Germany, 2008.
6. Calin, O.; Udriste, C. *Geometric Modeling in Probability and Statistics*; Springer International Publishing: Berlin, Germany, 2014.
7. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
8. Zhang, Z.; Sun, H.; Zhong, F. Information geometry of the power inverse Gaussian distribution. *Appl. Sci.* **2007**, *9*, 194–203.
9. Zhong, F.; Sun, H.; Zhang, Z. The geometry of the dirichlet manifold. *J. Korean Math. Soc.* **2008**. [[CrossRef](#)]
10. Amari, S. *Differential-Geometrical Methods in Statistics*; Springer: Berlin, Germany, 1985.
11. Takano, K. Exponential families admitting almost complex structures. *SUT J. Math.* **2010**, *46*, 1–21.
12. Nielsen, F. An information-geometric characterization of Chernoff information. *IEEE SPL* **2013**, *20*, 269–272. [[CrossRef](#)]
13. Ying, S.; Qin, H.; Peng, Y.; Wen, Z. Compute Karcher means on $SO(n)$ by the geometric conjugate gradient method. *Neurocomputing* **2016**, *215*, 169–174. [[CrossRef](#)]
14. Fiori, S.; Tanaka, T. An algorithm to compute averages on matrix Lie groups. *IEEE Trans. Signal Process.* **2010**, *57*, 4734–4743. [[CrossRef](#)]
15. Amari, S.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci.* **2010**, *58*, 183–195. [[CrossRef](#)]
16. Ay, N.; Amari, S. A novel approach to canonical divergence within information geometry. *Entropy* **2015**, *17*, 8111–8129. [[CrossRef](#)]
17. Amari, S. *Methods of Information Geometry*; Oxford University Press: Oxford, UK, 2000.
18. Kass, R.E.; Vos, P.W. *Geom. Found. Asymptot. Inference*; Wiley: Hoboken, NJ, USA, 1997.
19. Lee, J.M. *Riemannian Manifolds*; Springer: Berlin, Germany, 1997.
20. Kurose, T. Dual connections and affine geometry. *Math. Z.* **1990**, *203*, 115–121. [[CrossRef](#)]
21. Nielsen, F. Hypothesis testing, information divergence and computational geometry. *Geom. Sci. Inf.* **2013**. [[CrossRef](#)]
22. Kendall, W.S. Probability, convexity and harmonic maps with small image I: uniqueness and fine existence. *Proc. Lond. Math. Soc.* **1990**, *61*, 371–406. [[CrossRef](#)]
23. Buss, S.R.; Fillmore, J.P. Spherical averages and applications to spherical splines and interpolation. *ACM Trans. Graph.* **2001**, *20*, 95–126. [[CrossRef](#)]

