Article

Information-Geometric Approach for a One-Sided Truncated Exponential Family

Masaki Yoshioka [1,*] and **Fuyuhiko Tanaka** [2]

1   Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan
2   Center for Education in Liberal Arts and Sciences, Osaka University, Osaka 560-0043, Japan;
     ftanaka.celas@osaka-u.ac.jp
*   Correspondence: yoshioka@sigmath.es.osaka-u.ac.jp

Abstract: In information geometry, there has been extensive research on the deep connections between differential geometric structures, such as the Fisher metric and the α-connection, and the statistical theory for statistical models satisfying regularity conditions. However, the study of information geometry for non-regular statistical models is insufficient, and a one-sided truncated exponential family (oTEF) is one example of these models. In this paper, based on the asymptotic properties of maximum likelihood estimators, we provide a Riemannian metric for the oTEF. Furthermore, we demonstrate that the oTEF has an $\alpha = 1$ parallel prior distribution and that the scalar curvature of a certain submodel, including the Pareto family, is a negative constant.

1. Introduction

Information geometry is the study of the structure of statistical models using differential geometry. From the standpoint of geometry, a statistical model consisting of a collection of parameterized probability distributions can be regarded as a manifold. Then, when the statistical model satisfies a certain regularity condition, Chentsov's theorem leads to a natural differential geometric structure [1]. This natural differential geometric structure consists of the Riemann metric defined from the Fisher information matrix and a one-parameter family of affine connections, called the Fisher metric and the α-connection, respectively.

Information geometry of regular models has been studied for a long time, and deep relationships between the geometric structures and statistical models have been revealed (Amari [2] and Amari and Nagaoka [3]).

However, the geometric properties of statistical models which do not satisfy regularity conditions are not sufficiently investigated. One reason is that Chentsov's theorem cannot be applied to non-regular models; thus, a natural geometric structure is not established. For example, in regular models, it is possible to define the Fisher information matrix in two forms, but in statistical models where the support of the probability density function depends on the parameters, they do not coincide. In previous work, Amari [4] discussed the relationship between the Finsler geometry and non-regular models especially for the translation family.

In the present study, we discuss information geometry for a one-sided truncated exponential family (oTEF) [5], a typical non-regular model. An oTEF is a statistical model with two parameters: the natural parameter θ and the truncation parameter γ. The support of the probability density function depends on the truncation parameter γ, which is what makes such a model non-regular. However, similar to the exponential family, the derivatives of the log-likelihood function, moments, and KL divergence can be explicitly

described. Additionally, an oTEF model is a model containing many practical examples, such as the Pareto distribution family, truncated normal distribution family, and truncated exponential distribution family. The statistical properties of the oTEF have been studied for a long time, particularly in the area of parameter estimation theory (Bar-Lev [5], Akahira [6], and Akahira [7]). Recently, Nielsen [8] studied the relationship between the Kullback–Leibler divergence and the Bhattacharyya distance in the oTEF. Shemyakin [9] explored the Hellinger information matrix and its application to noninformative priors for multiparameter non-regular cases, which includes the oTEF, within the framework of Bayesian statistics.

The family of Pareto distributions, a type of oTEF, is a topic of study in information geometry as well (Rylov [10], Peng et al. [11], Li et al. [12], and Sun et al. [13]). As a result, it is known that the Pareto model has constant curvature with respect to the Levi-Civita connection. However, it is impossible to introduce a natural geometric structure, so the geometric structure of the Pareto distribution family has been formally defined based on regular models. Consequently, the consistency with statistical theory remains to be determined. Moreover, multiple geometric structures exist depending on the regular model expressions used as references. For example, Peng et al. [11], Rylov [10], and Sun et al. [13] all define the geometric structure of the Pareto distribution family based on the Fisher metric and α-connection, but these structures are different. Therefore, the question of which geometric structure should be used remains unresolved.

The results of the present study consist of two theorems: one regarding the α-parallel prior distribution and another concerning curvature. First of all, by comparing the relationship between the Fisher metric and the asymptotic behavior of the maximum likelihood estimators in regular models with those in the oTEF, we define the Riemannian metric for the oTEF. Additionally, we use a formal definition for the α-connection, similar to Sun et al. [13]. Under this geometric structure, we demonstrate two geometric properties. The first property is that the oTEF always has an α-parallel prior distribution with $\alpha = 1$. The α-parallel prior distribution is a type of noninformative prior distribution used in Bayesian statistics and is an information-geometric extension of Jeffreys prior distribution. Although the α-parallel prior distribution with $\alpha = 1$ always exists, this is not true for other α values. Such a phenomenon does not occur in regular models, highlighting the non-regularity. The second research result is related to the curvature of a submodel of the oTEF. We consider a submodel of the oTEF that includes common scale Pareto distribution families [14] and common location exponential distribution families [15] and demonstrate that its scalar curvature is constant. This result, stating that the scalar curvature of this submodel is constant, corresponds to a higher-dimensional version of the fact that the Pareto model has constant curvature [10].

The remainder of this paper is organized as follows. In Section 2, we review the geometric structure of regular statistical models and one-sided truncated exponential families, providing an overview of the relevant background. In Section 3, we provide a Riemannian metric and affine connection in the oTEF, giving the family a differential geometric structure. Under the geometric structure, we study the α-parallel prior to the oTEF in Section 4. Finally, Section 5 gives our derivation of the curvature for a submodel of the oTEF.

2. Preliminaries

2.1. Statistical Manifold

In this section, we review basic definitions and notation of information geometry for regular statistical models. See work by Amari [2] for more details.

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of distributions P_θ on a sample space χ parametrized by $\theta = (\theta^1, \ldots, \theta^n)$, where Θ is an open subset of $\mathbb{R}^n$. We treat the statistical model $\mathcal{P}$ as an n-dimensional manifold. In this case, parameter θ takes the role of coordinates of the manifold $\mathcal{P}$. Furthermore, we assume that parameter θ and distribution P_θ have a one-to-one correspondence. Furthermore, suppose that all distributions in $\mathcal{P}$ are absolutely

continuous with respect to a dominating measure μ. We denote the probability density function of distribution P_θ with respect to μ as $p(x, \theta)$.

In information geometry, we assume certain conditions on $\mathcal{P}$.

Definition 1 (Regular statistical models)**.** *A statistical model $\mathcal{P}$ is said to be regular if it satisfies the following conditions.*

1. *The support of the density function*

$$\operatorname{supp} p = \overline{\{x \in \chi : p(x, \theta) > 0\}} \tag{1}$$

   is independent of $\theta \in \Theta$.
2. *For all $x \in \chi$, the function $\theta \mapsto p(x, \theta)$ is C^∞ on Θ.*
3. *Let $l(x, \theta) = \log p(x, \theta)$. The elements of $\{\partial_1 l, \dots, \partial_n l\}$ are linearly independent as functions on χ.*
4. *For any functions in the following, partial differentiation ∂_i and integration with respect to the measure μ are interchangeable as*

$$\int_\chi \partial_i f(x, \theta) d\mu(x) = \partial_i \int_\chi f(x, \theta) d\mu(x). \tag{2}$$

The above four conditions are referred to as *regularity conditions*.

We will treat non-regular statistical models in later sections. In preparation, this subsection discusses the geometric structure of regular statistical models.

Next, we introduce the Fisher metric and α-connection. Let $\mathcal{P}$ be a regular statistical model and X be a random variable distributed according to P_θ. We usually use the following definition of the Riemannian metric and affine connections on $\mathcal{P}$.

Definition 2 (Fisher metric)**.** *The Fisher metric is the Riemannian metric on $\mathcal{P}$ defined by*

$$g_{ij}^F = \boldsymbol{E}_\theta \left[\partial_i l(X, \theta) \partial_j l(X, \theta) \right], \tag{3}$$

for $i, j = 1, \dots, n$. Here, the symbol $\boldsymbol{E}[\cdot]$ denotes the expectation with respect to the observation X.

Definition 3 (α-connection)**.** *The α-connection is an affine connection defined by the coefficients*

$$\overset{(\alpha)}{\Gamma}_{ij,k} = \boldsymbol{E} \left[\partial_i \partial_j l(X, \theta) \partial_k l(X, \theta) \right] + \frac{1 - \alpha}{2} \boldsymbol{E} \left[\partial_i l(X, \theta) \partial_j l(X, \theta) \partial_k l(X, \theta) \right], \tag{4}$$

for $i, j, k = 1, \dots, n$ and $\alpha \in \mathbb{R}$ on the coordinate system θ.

If $\mathcal{P}$ satisfies the regularity conditions, the Fisher metric and the α-connection have several properties. For instance, there is a formula [16] convenient for calculating the Fisher metric:

$$g_{ij}^F = -\mathbf{E} \left[\partial_i \partial_j l(x, \theta) \right]. \tag{5}$$

Furthermore, the α-connection and the $-\alpha$-connection are *dual*. In other words,

$$A\langle B, C \rangle = \left\langle \overset{(\alpha)}{\nabla}_A B, C \right\rangle + \left\langle B, \overset{(-\alpha)}{\nabla}_A C \right\rangle \tag{6}$$

holds for all tangent vectors A, B, and C, where $\langle \cdot, \cdot \rangle$ denotes the inner product given by the Fisher metric. In particular, the 0-connection is self-dual. Furthermore, the 0-connection corresponds to the *Levi-Civita connection*, defined by the coefficients

$$\overset{g}{\Gamma}_{ij,k} = \frac{1}{2} \left(\partial_i g^F_{jk} + \partial_j g^F_{ki} - \partial_k g^F_{ij} \right). \tag{7}$$

2.2. Prior Distributions and Volume Elements

Next, we introduce volume elements on a regular statistical model and its relation to Bayesian statistics [17].

In Bayesian statistics, for a given statistical model $\mathcal{P}$, we need a probability distribution over the model parameter space, which is called a *prior distribution* or simply a *prior*. We often denote a prior density as π. ($\pi(\theta) \geq 0$ and $\int_\Theta \pi(\theta) d\theta = 1$.)

A volume element on an n-dimensional oriented model manifold corresponds to a prior density function over the parameter space ($\theta \in \Theta \subset \mathbb{R}^n$) in a one-to-one manner. For a prior $\pi(\theta)$, its corresponding volume element ω is an n-form (differential form of degree n) and is written as

$$\omega = \pi(\theta) d\theta^1 \wedge \cdots \wedge d\theta^n \tag{8}$$

in the local coordinates.

For example, in the two-dimensional Euclidian space ($n = 2$), the volume element is given by $\omega = dx \wedge dy$ in the Cartesian coordinates (x, y). In the polar coordinates (r, ψ), it is written as $\omega = r dr \wedge d\psi$.

Now, let us explain noninformative priors (see, e.g., work by Robert [18] for details). If we have specific information on the parameter in advance, then the prior should reflect it, in which case it is often called a *subjective prior*. If not, we adopt a certain criterion and use a prior obtained through the criterion. Such priors are called *noninformative priors* or *objective priors*. In particular, the Jeffreys prior, which is given by $\sqrt{\det(g)}$, is the standard noninformative prior [19].

In information geometry, the Jeffreys prior is regarded as a 0-parallel volume element and was extended to an α-parallel volume element by Takeuchi and Amari [17]. The extensions are called α-parallel priors. They showed that the geometrical properties of regular models are deeply related to the existence of α-parallel priors.

When a suitable geometric structure is defined on a non-regular statistical model, it is interesting to see the relationship between the geometrical properties and the existence of α-parallel priors. Later sections will discuss this topic. In preparation, in this subsection, we briefly summarize some definitions and facts of α-parallel priors for regular models.

To define α-parallel priors, we introduce a geometric property of affine connections.

Definition 4 (equiaffine). *Let $\mathcal{P}$ be an n-dimensional manifold with an affine connection induced by a covariant derivative ∇.*

An affine connection ∇ is equiaffine if there exists a volume element ω such that

$$\nabla \omega = 0 \tag{9}$$

holds everywhere in $\mathcal{P}$. Furthermore, such a volume element ω is said to be a parallel volume element with respect to ∇.

The necessary and sufficient condition for an affine connection ∇ to be equiaffine is described by its curvature. The following proposition holds for a manifold with an affine connection ∇. Let $R_{ijk}{}^l$ be the components of the Riemannian curvature tensor [3] of ∇, defined as

$$R_{ijk}{}^l = \partial_i \Gamma_{jk}{}^l - \partial_j \Gamma_{ik}{}^l + \Gamma_{im}{}^l \Gamma_{jk}{}^m - \Gamma_{jm}{}^l \Gamma_{ik}{}^m, \quad (i, j, k, l = 1, \ldots, n) \tag{10}$$

where $\Gamma_{ij}{}^k$ denotes the connection coefficients of ∇.

Proposition 1 (Nomizu and Sasaki [20]). *The following conditions are equivalent:*

- ∇ *is equiaffine,*
- $R_{ijk}{}^k = 0,$

where $R_{ijk}{}^k = \sum_{k=1}^{n} R_{ijk}{}^k$.

The Levi-Civita connection is always equiaffine (Nomizu and Sasaki [20]).

Returning to the statistical model, we define the α-parallel prior distribution. Let $\mathcal{P}$ be an n-dimensional regular statistical manifold. It is shown that $\overset{(\alpha)}{\nabla}$ is equiaffine for some $\alpha \in \mathbb{R} \setminus \{0\}$ if and only if it is equiaffine for all $\alpha \in \mathbb{R}$. Such a statistical manifold $\mathcal{P}$ is said to be *statistically equiaffine*. In a statistically equiaffine manifold, we may represent the α-parallel volume element $\omega^{(\alpha)}$ as

$$\omega^{(\alpha)} = \pi^{(\alpha)}(\theta)d\theta^1 \wedge \cdots \wedge d\theta^n \tag{11}$$

for the coordinates θ, where $\pi^{(\alpha)} \in C^{\infty}(\mathcal{P})$. We take $\pi^{(\alpha)}(\theta)$ as a prior distribution on the parameter space Θ.

Definition 5 (α-parallel prior). *In a statistically equiaffine manifold, for fixed $\alpha \in \mathbb{R}$, we call the above form of $\pi^{(\alpha)}$ an α-parallel prior.*

Since the 0-connection (Levi-Civita connection) is equiaffine, there always exists a 0-parallel prior, known as the Jeffreys prior.

Takeuchi and Amari [17] give a necessary and sufficient condition for α-parallel priors to exist.

Proposition 2 (Takeuchi and Amari [17]). *For a model manifold $\mathcal{P}$, if*

$$\partial_i T_{jk}{}^k - \partial_j T_{ik}{}^k = 0 \quad (i, j = 1, \ldots, n), \tag{12}$$

then the α-parallel prior exists for any $\alpha \in \mathbb{R}$. Otherwise, only the 0-parallel prior exists.

2.3. One-Sided Truncated Exponential Family

In this section, we introduce a one-sided truncated exponential family, which is a typical non-regular model.

Definition 6 (One-sided Truncated Exponential Family (Bar-Lev [5])). *A one-sided truncated exponential family of distributions (oTEF) is a family $\mathcal{P} = \left\{ P_{\theta,\gamma} : \theta \in \Theta, \gamma \in (I_1, I_2) \right\}$ of distributions $P_{\theta,\gamma}$ with the density functions*

$$p(x, \theta, \gamma) = \exp\left\{ \sum_{i=1}^{n} \theta^i F_i(x) + C(x) - \psi(\theta, \gamma) \right\} \cdot \mathbb{1}_{[\gamma, I_2]}(x) \quad (x \in (I_1, I_2)) \tag{13}$$

with respect to the Lebesgue measure. Here, $-\infty \leq I_1 < I_2 \leq \infty$ are known parameters, C is a continuous function, and $F_1, \ldots, F_n$ are absolutely continuous with $dF_1(x)/dx, \ldots, dF_n(x)/dx \neq 0$ over the interval (I_1, I_2).

We say $\theta = (\theta^1, \ldots, \theta^n)$ is the natural parameter and γ is the truncation parameter. We also denote the interval (I_1, I_2) as I.

In oTEF $\mathcal{P}$, the support of the density function depends on the truncation parameter γ. Then, the oTEF does not satisfy the regularity conditions. An oTEF differs from an exponential family at this point. The density of an exponential family has a support that is independent of the parameters.

Furthermore, this truncation parameter γ does not allow for the interchange of partial differentiation ∂_γ and integration with respect to the Lebesgue measure. For instance, this means

$$\mathbf{E}[\partial_\gamma l(X, \theta, \gamma)] = -\partial_\gamma \psi(\theta, \gamma) \tag{14}$$

instead of $\mathbf{E}[\partial_\gamma l(X, \theta, \gamma)] = 0$.

Here, we introduce two properties of the function ψ.

First, the partial derivative $-\partial_\gamma \psi(\theta, \gamma)$ coincides with $p(x, \theta, \gamma)|_{x=\gamma}$ and is always positive. This fact can be verified as follows. Since p is a probability density function,

$$\exp\{\psi(\theta, \gamma)\} = \int_\gamma^{I_2} \exp\left\{\sum_{i=1}^n \theta^i F_i(x) + C(x)\right\} dx. \tag{15}$$

Therefore, by differentiating both sides with respect to γ, we obtain

$$\partial_\gamma \psi(\theta, \gamma) \exp\{\psi(\theta, \gamma)\} = -\exp\left\{\sum_{i=1}^n \theta^i F_i(\gamma) + C(\gamma)\right\} \tag{16}$$

and

$$-\partial_\gamma \psi(\theta, \gamma) = p(x, \theta, \gamma)|_{x=\gamma} > 0. \tag{17}$$

Second, the following lemma holds.

Lemma 1. *For the function $\psi(\theta, \gamma)$ in* (13)*, there always exists θ and γ such that $\partial_i \partial_\gamma \psi(\theta, \gamma) \neq 0$.*

Proof. Suppose that $\partial_\gamma \partial_i \psi(\theta, \gamma)$ is always 0. Under this assumption, we will show that $\partial_i p(x, \theta, \gamma) \equiv 0$, leading to a contradiction that θ is not a parameter.

If $\partial_\gamma \partial_i \psi(\theta, \gamma) \equiv 0$, the function $\psi(\theta, \gamma)$ can be expressed as

$$\psi(\theta, \gamma) = \psi_1(\theta) + \psi_2(\gamma), \tag{18}$$

where $\psi_1 \in C^\infty(\Theta)$, $\psi_2 \in C^\infty(I)$.

Then,

$$\exp\{\psi_1(\theta) + \psi_2(\gamma)\} = \int_\gamma^{I_2} \exp\left\{C(x) + \sum \theta^i F_i(x)\right\} dx. \tag{19}$$

Differentiating both sides with respect to γ, we obtain

$$\psi_2'(\gamma) e^{\psi_1(\theta) + \psi_2(\gamma)} = -\exp\left\{C(\gamma) + \sum \theta^i F_i(\gamma)\right\}, \tag{20}$$

$$\log\left\{-\psi_2'(\gamma) e^{\psi_2(\gamma) - C(\gamma)}\right\} + \psi_1(\theta) = \sum \theta^i F_i(\gamma). \tag{21}$$

Further differentiating both sides with respect to θ^i, we have

$$\partial_i \psi_1(\theta) = F_i(\gamma) \tag{22}$$

and all F_i are constants.

Therefore,

$$\partial_i p(x, \theta, \gamma) = p(x, \theta, \gamma) \partial_i \left\{C(x) + \sum \theta^i F_i - \psi_1(\theta) - \psi_2(\gamma)\right\} \tag{23}$$

$$= p(x, \theta, \gamma)\{F_i - \partial_i \psi_1(\theta)\} \tag{24}$$

$$= 0. \tag{25}$$

□

The family of Pareto distributions is an example of the oTEF. Pareto distributions have the following density functions,

$$p(x, \theta, \gamma) = \frac{\theta \gamma^\theta}{x^{\theta+1}} \cdot \mathbb{1}_{[\gamma, \infty)}(x) \tag{26}$$

$$= \exp(-(1 + \theta) \log x + \log \theta + \theta \log \gamma) \cdot \mathbb{1}_{[\gamma, \infty)}(x) \tag{27}$$

with the natural parameter $\theta \in (0, \infty)$ and the truncation parameter $\gamma \in (0, \infty)$. This family is used to describe various natural and social phenomena [21].

To discuss geometric structures in subsequent sections, we focus on the asymptotic behavior of maximum likelihood estimators in the oTEF. Our discussion of these estimators follows previous works. Consider random variables $X_1, \ldots, X_N$ independent and identically distributed according to $P_{\theta, \gamma}$, and let $X_{(1)} \leq, \cdots \leq, X_{(N)}$ be the order statistics of the sample. Let $\hat{\theta}$ and $\hat{\gamma}$ denote the maximum likelihood estimators for θ and γ, respectively. Bar-Lev [5] showed the existence and uniqueness of $\hat{\theta}$ and $\hat{\gamma}$. Here, $\hat{\gamma} = X_{(1)}$ and $\hat{\theta}$ is a root of the maximum likelihood equation $\partial_i l(X, \hat{\theta}, \hat{\gamma}) = 0$ for $i = 1, \ldots, n$.

The first-order asymptotic variances of $\hat{\theta}$ and $\hat{\gamma}$ are given by

$$\mathbf{V}[\hat{\theta}] = \frac{1}{N \partial_i \partial_j \psi(\theta, \gamma)} + \mathrm{O}\left(\frac{1}{N^2}\right), \tag{28}$$

$$\mathbf{V}[\hat{\gamma}] = \frac{1}{N^2 \{\partial_\gamma \psi(\theta, \gamma)\}^2} + \mathrm{O}\left(\frac{1}{N^3}\right), \tag{29}$$

$$\mathbf{Cov}[\hat{\theta}, \hat{\gamma}] = \mathrm{O}\left(\frac{1}{N^3}\right). \tag{30}$$

These are essential to our argument in the next section. Additionally, Akahira [6] and Akahira and Ohyauchi [22] obtained the second-order asymptotic loss.

3. Geometric Structure on One-Sided Exponential Families

This section gives definitions of a geometric structure of the oTEF. We take $\mathcal{P}$ as an $(n + 1)$-dimensional manifold with coordinates $(\theta^1, \ldots, \theta^n, \gamma)$. Since $\mathcal{P}$ does not satisfy regularity conditions, its geometric structure is not determined in a natural way.

3.1. Riemannian Metric in the oTEF

In this subsection, we define a Riemannian metric on oTEF $\mathcal{P}$ as follows.

Definition 7 (Riemannian metric in the oTEF). *Let* $\mathcal{P} = \{P_{\theta, \gamma} : \theta \in \Theta, \gamma \in I\}$ *belong to the oTEF. The Riemannian metric of the oTEF is defined by*

$$g_{ij} = \mathbf{E}\left[\partial_i l(x, \theta, \gamma) \partial_j l(x, \theta, \gamma)\right], \tag{31}$$

$$g_{i\gamma} = 0, \tag{32}$$

$$g_{\gamma\gamma} = \{\partial_\gamma \psi(\theta, \gamma)\}^2 \tag{33}$$

for $i, j = 1, \ldots, n$. *This metric is also represented as*

$$(g) = \begin{pmatrix} g_{11} & \cdots & g_{1n} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ g_{n1} & \cdots & g_{nn} & 0 \\ 0 & \cdots & 0 & \{\partial_\gamma \psi(\theta, \gamma)\}^2 \end{pmatrix}. \tag{34}$$

Note that g_{ij} can be expressed in terms of the function ψ as follows:

$$g_{ij} = -\mathbf{E}\left[\partial_i\partial_j l(x,\theta,\gamma)\right] = \partial_i\partial_j\psi(\theta,\gamma). \tag{35}$$

This is similar to the case of exponential family distributions.

We will now explain how we came to the above definition.

Consider an exponential family $\mathcal{E}_\gamma = \left\{P_{\theta,\gamma} : \theta \in \Theta\right\}$ for $\gamma \in I$ and a statistical model $\mathcal{F}_\theta = \left\{P_{\theta,\gamma} : \gamma \in I\right\}$ for $\theta \in \Theta$. $\mathcal{E}_\gamma$ is an n-dimensional submanifold of $\mathcal{P}$ obtained by fixing the truncation parameter γ and satisfies the regularity conditions. Additionally, $\mathcal{F}_\theta$ is a one-dimensional submanifold of $\mathcal{P}$ obtained by fixing the natural parameter θ. Since $\mathcal{E}_\gamma$ is regular, the Riemannian metric on $\mathcal{E}_\gamma$ should be the Fisher metric defined in Definition 2. This idea induces the components g_{ij} to be the components of the Fisher metric. The remaining task is to define $g_{i\gamma}$, the inner products of ∂_i and ∂_γ, and $g_{\gamma\gamma}$, the metric on $\mathcal{F}_\theta$.

In this context, we review the statistical interpretation of the Fisher metric in regular models. Let $\mathcal{P}_0$ be a regular model with parameters $\theta \in \mathbb{R}^n$. As mentioned in Section 2.1, the Fisher metric is a Riemannian metric defined by the Fisher information matrix. Expanding the variance of the maximum likelihood estimator $\hat{\theta}$, we have

$$\mathbf{V}\left[\hat{\theta}\right] = \frac{1}{N}\left(g_{ij}^F\right)^{-1} + \mathrm{O}\left(\frac{1}{N^2}\right), \tag{36}$$

where the first-order coefficient corresponds to $\left(g_{ij}^F\right)^{-1}$.

On the other hand, in the oTEF, the Riemannian metric is determined from the coefficient of the variance of the maximum likelihood estimator. As shown in Section 2.3, the variances of $\hat{\theta}$ and $\hat{\gamma}$ are expressed as

$$\mathbf{V}\left[\hat{\theta}\right] = \frac{1}{N}\left(g_{ij}\right)^{-1} + \mathrm{O}\left(\frac{1}{N^2}\right), \tag{37}$$

$$\mathbf{V}[\hat{\gamma}] = \frac{1}{N^2\{\partial_\gamma\psi(\theta,\gamma)\}^2} + \mathrm{O}\left(\frac{1}{N^3}\right), \tag{38}$$

$$\mathbf{Cov}\left[\hat{\theta},\hat{\gamma}\right] = \mathrm{O}\left(\frac{1}{N^3}\right), \tag{39}$$

where (g_{ij}) is the matrix $(g_{ij})_{i,j=1}^n$. Then, similarly to the Fisher information matrix, $\{\partial_\gamma\psi(\theta,\gamma)\}^2$ appears as the reciprocal of the first-term coefficient. From this, the Riemannian metric of $\mathcal{F}_\theta$ is defined as

$$g_{\gamma\gamma} = \{\partial_\gamma\psi(\theta,\gamma)\}^2. \tag{40}$$

Furthermore, it should be noted that $\mathbf{Cov}\left[\hat{\theta},\hat{\gamma}\right]$ is negligible up to $\mathrm{O}\left(\frac{1}{N^2}\right)$. Moreover, whether θ is known does not affect the first-order term of $\mathbf{V}\left[\hat{\theta}\right]$ [22]. This is also true for the estimation of θ [23]. Based on these facts, we assume that ∂_i and ∂_γ are orthogonal, and define

$$g_{i\gamma} = 0 \quad (i = 1,\ldots,n). \tag{41}$$

As a result, the Riemannian metric in Definition 7 is obtained.

This Riemannian metric is equal to the formally defined Fisher metric on the oTEF. In other words, the equality

$$g_{ab} = \mathbf{E}[\partial_a l(X,\theta,\gamma)\partial_b l(X,\theta,\gamma)] \tag{42}$$

holds for $a, b = 1,\ldots,n,\gamma$. The right-hand side is the same as the definition of the Fisher metric on regular statistical models.

However, the Riemann metric does not satisfy the equation

$$g_{a\gamma} = -\mathbf{E}[\partial_a \partial_\gamma l(x, \theta, \gamma)] \quad (a = 1, \ldots, n, \gamma). \tag{43}$$

For $i = 1, \ldots, n$, we have

$$g_{i\gamma} = 0, \tag{44}$$

$$\mathbf{E}[\partial_i \partial_\gamma l(x, \theta, \gamma)] = \partial_i \partial_\gamma \psi(\theta, \gamma), \tag{45}$$

but $\partial_i \partial_\gamma \psi(\theta, \gamma) \neq 0$ for some θ and γ, by Lemma 1. This is influenced by the fact that it does not satisfy the regularity conditions.

3.2. Affine Connections in the oTEF

Next, we define an affine connection in the oTEF. In this study, we adopt two types of connections: the Levi-Civita connection and the α-connection.

The first affine connection is the *Levi-Civita connection*:

$$\overset{g}{\Gamma}_{ab,c} = \partial_a g_{bc} + \partial_b g_{ca} - \partial_c g_{ab} \tag{46}$$

introduced from the Riemannian metric. Of course, this Levi-Civita connection is also a metric and self-dual, the same as in the regular case. Rylov [10] and Li et al. [12] previously adopted this same affine connection.

Second, we define an affine connection in the oTEF as an analogy for the α-connection in the regular statistical model defined in Definition 3.

Definition 8. *For a given $\alpha \in \mathbb{R}$, the α-connection in the oTEF is defined by the connection coefficients*

$$\overset{(\alpha)}{\Gamma}_{ab,c}(\theta, \gamma) = \mathbf{E}[\partial_a \partial_b l \partial_c l] + \frac{1 - \alpha}{2} \mathbf{E}[\partial_a l \partial_b l \partial_c l] \quad (a, b, c = 1, \ldots, n, \gamma), \tag{47}$$

where $l = l(X, \theta, \gamma)$ is a log-likelihood function.

The above definition is obtained by substituting the oTEF probability density function into Equation (4) for the α-connection in the regular model. In particular, for the Pareto distribution family, it coincides with the α-connection given by Sun et al. [13] (see the equation in their paper). Note that, the log-likelihood function is not differentiable with respect to γ at $\gamma = x$. We calculate the expectation over the interval (γ, I_2) instead of $[\gamma, I_2)$.

This α-connection is torsion-free,

$$\overset{(\alpha)}{\Gamma}_{ab,c} = \overset{(\alpha)}{\Gamma}_{ba,c}. \tag{48}$$

The connection coefficients for the α-connection are given by

$$\overset{(\alpha)}{\Gamma}_{ij,k} = \frac{1-\alpha}{2}\partial_i\partial_j\partial_k\psi, \tag{49}$$

$$\overset{(\alpha)}{\Gamma}_{ij,\gamma} = \frac{1+\alpha}{2}\partial_i\partial_j\psi\partial_\gamma\psi, \tag{50}$$

$$\overset{(\alpha)}{\Gamma}_{i\gamma,j} = -\frac{1-\alpha}{2}\partial_i\partial_j\psi\partial_\gamma\psi, \tag{51}$$

$$\overset{(\alpha)}{\Gamma}_{i\gamma,\gamma} = \partial_i\partial_\gamma\psi\partial_\gamma\psi, \tag{52}$$

$$\overset{(\alpha)}{\Gamma}_{\gamma\gamma,i} = 0, \tag{53}$$

$$\overset{(\alpha)}{\Gamma}_{\gamma\gamma,\gamma} = \partial_\gamma\partial_\gamma\psi\partial_\gamma\psi - \frac{1-\alpha}{2}(\partial_\gamma\psi)^3, \tag{54}$$

where $\psi = \psi(\theta, \gamma)$.

The α-connection in the oTEF differs from the one in the regular model in several aspects.

First, the 0-connection does not correspond to the Levi-Civita connection. This is due to the inability to express the α-connection coefficients in the following form:

$$\overset{(\alpha)}{\Gamma}_{abc}(\theta, \gamma) = \overset{g}{\Gamma}_{abc}(\theta, \gamma) - \frac{\alpha}{2}\mathbf{E}[\partial_a l\partial_b l\partial_c l]. \tag{55}$$

This transformation involves an interchange of the order of differentiation and integration, which is one of the regularity conditions. Therefore, in the non-regular model oTEF, the two sides do not match.

Additionally, the dual connection of the α-connection in the oTEF does not become the $-\alpha$-connection. This can be verified as follows.

The partial derivative $\partial_\gamma g_{i\gamma}$ and connection coefficients $\overset{(\alpha)}{\Gamma}_{\gamma i,\gamma}$ and $\overset{(-\alpha)}{\Gamma}_{\gamma\gamma,i}$ are given by

$$\partial_\gamma g_{\gamma i} = 0, \tag{56}$$

$$\overset{(\alpha)}{\Gamma}_{\gamma i,\gamma} = \partial_i\partial_\gamma\psi(\theta, \gamma)\partial_\gamma\psi(\theta, \gamma), \tag{57}$$

$$\overset{(-\alpha)}{\Gamma}_{\gamma\gamma,i} = 0. \tag{58}$$

Thus,

$$\overset{(\alpha)}{\Gamma}_{\gamma i,\gamma} + \overset{(-\alpha)}{\Gamma}_{\gamma\gamma,i} = \partial_\gamma\partial_i\psi(\theta, \gamma)\partial_\gamma\psi(\theta, \gamma). \tag{59}$$

Therefore, since $\partial_\gamma\partial_i\psi(\theta, \gamma) \neq 0$ and $\partial_\gamma\psi(\theta, \gamma) < 0$, the duality does not hold.

4. Existence of an α-Parallel Prior in the oTEF

In this section, as part of investigating the properties of the α-connection in the oTEF, we deal with the α-parallel priors. We show that there exists an α-parallel prior distribution for $\alpha = 1$.

The existence of α-parallel priors depends on the geometric properties of statistical models, and they are not guaranteed to exist in general. Therefore, it is necessary to investigate the existence of α-parallel priors in the case of the oTEF. Note that the Levi-Civita connection is always equiaffine, and it is known to have the Jeffreys prior as a parallel prior distribution. Therefore, we do not deal with it in this section.

In the oTEF, attention is needed to be paid to the conditions for the existence of the α-parallel prior distributions. In the case of regular models, Proposition 2 provides a necessary and sufficient condition for the existence of α-parallel priors. Sun et al. [13] revealed that

the Pareto distribution family does not satisfy this condition and claims that it does not have α-parallel priors. However, the above deduction is incorrect, since Proposition 2 does not hold for an oTEF distribution. In proof [17] of Proposition 2, the connection coefficients of the α-connection are written in the same form as (55) by the Levi-Civita connection and the cubic tensor $T_{ijk} = \mathbf{E}\left[\partial_i l \partial_j l \partial_k l\right]$. However, this form cannot represent the α-connection in Definition 8. Then, in the oTEF, Proposition 2 does not confirm that the equation

$$\partial_a T_{bc}{}^c - \partial_b T_{ac}{}^c = 0 \quad (a, b = 1, \dots, n, \gamma) \tag{60}$$

is a necessary and sufficient condition for the existence of α-parallel priors.

Therefore, in this study, we use the necessary and sufficient conditions for general affine connections to investigate the existence of the α-parallel prior.

The following theorem reveals the existence of α-parallel priors in the oTEF.

Theorem 1. *Consider $\mathcal{P}$ as belonging to the oTEF with densities of the form as those in Definition 13 with the natural parameter $\theta \in \Theta$ and the truncation parameter γ. If $\alpha = 1$, then the connection $\overset{(\alpha)}{\nabla}$ is equiaffine and there exists a one-parallel prior. Moreover, this one-parallel prior $\pi^{(1)}$ can be represented as*

$$\pi^{(1)}(\theta, \gamma) \propto -\partial_\gamma \psi(\theta, \gamma) \tag{61}$$

for $\theta \in \Theta$, $\gamma \in I$.

Proof. First, we prove that $\overset{(1)}{\nabla}$ is equiaffine. By Proposition 1, the necessary and sufficient condition for $\overset{(\alpha)}{\nabla}$ to be equiaffine is $\overset{(\alpha)}{R}_{abc}{}^c = 0$ for $a, b = 1, \dots, n, \gamma$ everywhere in $\mathcal{P}$, where $\overset{(\alpha)}{R}$ denotes α-Riemannian curvature tensors. This condition can be represented as $\partial_a \overset{(\alpha)}{\Gamma}_{bc}{}^c = \partial_b \overset{(\alpha)}{\Gamma}_{ac}{}^c$, since

$$\overset{(\alpha)}{R}_{abc}{}^c = \sum_{c,d=c} \left\{ \partial_a \overset{(\alpha)}{\Gamma}_{bc}{}^d - \partial_b \overset{(\alpha)}{\Gamma}_{ac}{}^d + \overset{(\alpha)}{\Gamma}_{ae}{}^d \overset{(\alpha)}{\Gamma}_{bc}{}^e - \overset{(\alpha)}{\Gamma}_{be}{}^d \overset{(\alpha)}{\Gamma}_{ac}{}^e \right\} \tag{62}$$

$$= \partial_a \overset{(\alpha)}{\Gamma}_{bc}{}^c - \partial_b \overset{(\alpha)}{\Gamma}_{ac}{}^c + \overset{(\alpha)}{\Gamma}_{ae}{}^c \overset{(\alpha)}{\Gamma}_{bc}{}^e - \overset{(\alpha)}{\Gamma}_{be}{}^c \overset{(\alpha)}{\Gamma}_{ac}{}^e \tag{63}$$

$$= \partial_a \overset{(\alpha)}{\Gamma}_{bc}{}^c - \partial_b \overset{(\alpha)}{\Gamma}_{ac}{}^c. \tag{64}$$

By Definition 7, we have

$$\overset{(\alpha)}{\Gamma}_{ja}{}^a = \frac{1-\alpha}{2} \partial_j \log(\det(g_{kl})) + \partial_j \log(-\partial_\gamma \psi), \tag{65}$$

$$\overset{(\alpha)}{\Gamma}_{\gamma a}{}^a = -\frac{(n+1)(1-\alpha)}{2} \partial_\gamma \psi + \partial_\gamma \log(-\partial_\gamma \psi). \tag{66}$$

Thus,

$$\partial_i \overset{(\alpha)}{\Gamma}_{ja}{}^a - \partial_j \overset{(\alpha)}{\Gamma}_{ia}{}^a = 0, \tag{67}$$

$$\partial_\gamma \overset{(\alpha)}{\Gamma}_{ia}{}^a - \partial_i \overset{(\alpha)}{\Gamma}_{\gamma a}{}^a = \frac{1-\alpha}{2} \partial_i \partial_\gamma \log(\det(g_{kl})) + (n+1)\partial_i \partial_\gamma \psi. \tag{68}$$

Therefore, the formula $\partial_a \overset{(\alpha)}{\Gamma}_{bc}{}^a = \partial_b \overset{(\alpha)}{\Gamma}_{ac}{}^c$ holds when $\alpha = 1$ and $\overset{(1)}{\nabla}$ is equiaffine.

Second, we find the one-parallel prior. Let $\pi^{(1)}$ be the density of the one-parallel volume element. According to Proposition 1 in the paper by Takeuchi and Amari [17], we have

$$\partial_a \log \pi^{(1)} = \overset{(1)}{\Gamma}_{ab}{}^b. \tag{69}$$

In the case of the oTEF, its representation is

$$\partial_i \log \pi^{(1)} = \partial_i \log(-\partial_\gamma \psi), \partial_\gamma \log \pi^{(1)} = \partial_\gamma \log(-\partial_\gamma \psi). \tag{70}$$

Therefore, the one-parallel prior for $\mathcal{P}$ is given as

$$\pi^{(1)}(\theta, \gamma) \propto -\partial_\gamma \psi. \tag{71}$$

$\square$

Moreover, the above one-parallel prior coincides with a certain reference prior. A reference prior, proposed by Bernardo [24], is a noninformative prior distribution derived from an information-theoretic perspective. Specifically, it is defined as a prior distribution that maximizes the expectation of the KL divergence between the posterior and prior distributions, and in the case of no nuisance parameters, it coincides with the Jeffreys prior [24]. However, this is not necessarily the case when nuisance parameters are present. Ghosh and Mukerjee [25] provided a new formulation for reference priors with nuisance parameters by considering the maximization of a functional with an appropriate penalty term. Furthermore, Ghosal [26] extended this reference prior to non-regular models where the support of the density depends on the parameters. When applied to an oTEF model with γ as the parameter of interest and θ as the nuisance parameter, the reference prior of Ghosal [26] is given by

$$\pi_{\text{Ghosal}}(\theta, \gamma) \propto -\partial_\gamma \psi. \tag{72}$$

Thus, it coincides with the one-parallel prior in Theorem 1.

5. Scalar Curvature on a Submodel of the oTEF

This section finds a submodel of the oTEF with constant scalar curvature for the Levi-Civita connection. We do not use the α-connection in Definition 8.

Previous works about the information geometry of non-regular cases have mainly studied the geometric structure of Pareto distributions for the Levi-Civita connection. They adopted the formally defined Fisher metric as in (42), which is consistent with our Riemannian metric. Rylov [10] found that the family of Pareto distributions has a constant curvature with respect to the Levi-Civita connection. Li et al. [12] showed that its geometrical structure is isometric to the Poincaré upper half-plane and applied this geometrical structure to Bayesian inference by considering the Jeffreys prior.

We extend these previous works on Pareto distributions to n dimensions. For $i = 1, \ldots, n$, let F_i be a smooth function on the interval I and let X_i be a random variable following a two-parameter truncated exponential distribution, which has density function

$$p_E(x, \theta^i, \gamma) = \theta^i \exp\left\{-\theta^i(x - \gamma)\right\} \cdot \mathbb{1}_{[\gamma, \infty)}(x) \quad (x \in \mathbb{R}) \tag{73}$$

with respect to the Lebesgue measure, where $\theta^i \in (0, \infty)$, $\gamma \in \mathbb{R}$. An oTEF includes this distribution.

Consider $Q_{\theta,\gamma}$, a joint distribution of independent random variables $F_1(X_1),\ldots,F_n(X_n)$ with the common redefined truncation parameter γ and the natural parameter $\theta = (\theta^1,\ldots,\theta^n)$. $Q_{\theta,\gamma}$ has the density function

$$q(x,\theta,\gamma) = \exp\left\{ -\sum_{i=1}^n \theta^i\{F_i(x_i) - F_i(\gamma)\} + \sum_{i=1}^n \log\theta^i \right\} \cdot \mathbb{1}_{I^n(\gamma)}(x) \tag{74}$$

$$(x = (x_1,\ldots,x_n) \in I^n) \tag{75}$$

with respect to the Lebesgue measure on $\mathbb{R}^n$. Here, I^n is a rectangle $\prod_{i=1}^n F_i(\mathbb{R})$ and $I(\gamma)$ is $I^n \cap [\gamma,\infty)^n$. $\mathcal{Q}$ is a family of distributions $Q_{\theta,\gamma}$ with parameters θ and γ.

$\mathcal{Q}$ includes practical examples such as a family of several Pareto distributions with a common scale parameter (Rohatgi and Saleh [14]) and a family of truncated exponential distributions with a common location parameter (Ghosh and Razmpour [15]). Truncated exponential distributions with a common γ are sometimes applied to reliability and life testing. Please assume that the case where the first failure of n products can occur only after a common minimum time γ has elapsed, and these products have unknown and possibly unequal failure rates $\theta^1,\ldots,\theta^n$. This truncation parameter γ takes the role of a "guarantee time", so estimation of the parameter γ is vital to determining the warranty period.

Note that the geometric structure of the above common truncation parameter model is equivalent to that of the family of Pareto distributions when $n = 1$.

Theorem 2. *Riemannian manifold $\mathcal{Q}$ has a constant scalar curvature of -2.*

Proof. Let $G(\theta,\gamma)$ denote $\sum_i \theta^i F_i'(\gamma)$, where $F'(\gamma) = dF(\gamma)/d\gamma$.

The Riemannian metric for the common truncation parameter model is given by

$$g_{ij} = \frac{1}{(\theta^i)^2}\delta_{ij}, \tag{76}$$

$$g_{i\gamma} = 0, \tag{77}$$

$$g_{\gamma\gamma} = (G(\theta,\gamma))^2 \tag{78}$$

for $i,j = 1,\ldots,n$. The tangent vectors $\partial_1,\ldots,\partial_n,\partial_\gamma$ are mutually orthogonal. By (46), we have

$$\Gamma_{ij,k} = -\frac{1}{(\theta^i)^3}\delta_{ijk}, \tag{79}$$

$$\Gamma_{ij,\gamma} = 0, \tag{80}$$

$$\Gamma_{i\gamma,j} = 0, \tag{81}$$

$$\Gamma_{i\gamma,\gamma} = F_i'(\gamma)G(\theta,\gamma), \tag{82}$$

$$\Gamma_{\gamma\gamma,i} = -F_i'(\gamma)G(\theta,\gamma), \tag{83}$$

$$\Gamma_{\gamma\gamma,\gamma} = G(\theta,\gamma)\partial_\gamma G(\theta,\gamma), \tag{84}$$

and

$$\Gamma_{ij}{}^{k} = -\frac{1}{\theta^{i}}\delta_{ijk}, \tag{85}$$

$$\Gamma_{ij}{}^{\gamma} = 0, \tag{86}$$

$$\Gamma_{i\gamma}{}^{j} = 0, \tag{87}$$

$$\Gamma_{i\gamma}{}^{\gamma} = \frac{F_{i}'(\gamma)}{G(\theta,\gamma)} \tag{88}$$

$$= \partial_{i}\log G(\theta,\gamma), \tag{89}$$

$$\Gamma_{\gamma\gamma}{}^{i} = -\frac{F_{i}'(\gamma)}{G(\theta,\gamma)} \tag{90}$$

$$= -\partial_{i}\log G(\theta,\gamma), \tag{91}$$

$$\Gamma_{\gamma\gamma}{}^{\gamma} = \frac{\partial_{\gamma}G(\theta,\gamma)}{G(\theta,\gamma)} \tag{92}$$

$$= \partial_{\gamma}\log G(\theta,\gamma) \tag{93}$$

for $i,j,k = 1,\ldots,n$, where

$$\delta_{ijk} = \begin{cases} 1 & \text{for } i = j = k, \\ 0 & \text{otherwise.} \end{cases} \tag{94}$$

Hence, we obtain

$$R_{i\gamma i}{}^{\gamma} = \partial_{i}\,\Gamma_{\gamma i}{}^{\gamma} - \partial_{\gamma}\,\Gamma_{ii}{}^{\gamma} + \Gamma_{ia}{}^{\gamma}\,\Gamma_{\gamma i}{}^{a} - \Gamma_{\gamma a}{}^{\gamma}\,\Gamma_{ii}{}^{a} \tag{95}$$

$$= \partial_{i}\partial_{i}\log G(\theta,\gamma) + (\partial_{i}\log G(\theta,\gamma))^{2} + \frac{1}{\theta^{i}}\partial_{i}\log G(\theta,\gamma) \tag{96}$$

$$= -\left(\frac{F_{i}'(\gamma)}{G(\theta,\gamma)}\right)^{2} + \left(\frac{F_{i}'(\gamma)}{G(\theta,\gamma)}\right)^{2} + \frac{F_{i}'(\gamma)}{\theta^{i}G(\theta,\gamma)} \tag{97}$$

$$= \frac{F_{i}'(\gamma)}{\theta^{i}G(\theta,\gamma)} \tag{98}$$

for $i = 1,\ldots,n$.

Therefore, the scalar curvature is given by

$$\mathcal{R} = R_{abcd}g^{ad}g^{bc} \tag{99}$$

$$= -2. \tag{100}$$

□

6. Concluding Remarks

This paper considered the geometric structure of a one-sided truncated exponential family (oTEF) with parameters θ and γ. We constructed a Riemannian metric based on the asymptotic properties of the maximum likelihood estimators. Under this, we showed that the formally defined α-connection admits α-parallel priors when $\alpha = 1$. Our result gives geometric meaning to a specific reference prior. Furthermore, we proved that the scalar curvature of some submodels of the oTEF obtained by making γ common across multiple distributions is constant.

It is essential to discuss suitable affine conditions for the oTEF. First, we need to reveal the statistical meaning of the α-connection in the oTEF. α-connection coefficients are expected to appear in higher-order terms of the variance of maximum likelihood estimators. Additionally, instead of the α-connection, we can construct a family of affine connections

to be equiaffine by connecting the *α*-connection and the Levi-Civita connection. It is also interesting to consider affine connections induced by the third derivatives of divergences.

References

1. Chentsov, N.N. Statistical Decision Rules and Optimal Inference. In *Translations of Mathematical Monographs*; American Mathematical Society: Providence, RI, USA, 1982; Volume 53. [CrossRef]
2. Amari, S. Differential-Geometrical Methods in Statistics. In *Lecture Notes in Statistics*; Springer: Berlin, Germany, 1985; Volume 28. [CrossRef]
3. Amari, S.; Nagaoka, H. Methods of Information Geometry. In *Translations of Mathematical Monographs*; American Mathematical Society: Providence, RI, USA; Oxford University Press: Oxford, UK, 2000. [CrossRef]
4. Amari, S. Finsler geometry of non-regular statistical models. *RIMS Kokyuroku* **1984**, *538*, 81–95.
5. Bar-Lev, S.K. Large Sample Properties of the Mle and Mcle for the Natural Parameter of a Truncated Exponential Family. *Ann. Inst. Stat. Math.* **1984**, *36*, 217–222. [CrossRef]
6. Akahira, M. Second-Order Asymptotic Comparison of the MLE and MCLE of a Natural Parameter for a Truncated Exponential Family of Distributions. *Ann. Inst. Stat. Math.* **2016**, *68*, 469–490. [CrossRef]
7. Akahira, M. *Statistical Estimation for Truncated Exponential Families*; Springer Briefs in Statistics; Springer: Singapore, 2017. [CrossRef]
8. Nielsen, F. Statistical Divergences between Densities of Truncated Exponential Families with Nested Supports: Duo Bregman and Duo Jensen Divergences. *Entropy* **2022**, *24*, 421. [CrossRef] [PubMed]
9. Shemyakin, A. Hellinger Information Matrix and Hellinger Priors. *Entropy* **2023**, *25*, 344. [CrossRef] [PubMed]
10. Rylov, A. Constant Curvature Connections On Statistical Models. In *Information Geometry and Its Applications*; Ay, N., Gibilisco, P., Matúš, F., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 252, pp. 349–361. [CrossRef]
11. Peng, L.; Sun, H.; Jiu, L. The Geometric Structure of the Pareto Distribution. *Geom. Struct. Pareto Distrib.* **2007**, *14*, 5–13.
12. Li, M.; Sun, H.; Peng, L. Fisher–Rao Geometry and Jeffreys Prior for Pareto Distribution. *Commun. Stat.—Theory Methods* **2022**, *51*, 1895–1910. [CrossRef]
13. Sun, F.; Cao, Y.; Zhang, S.; Sun, H. The Bayesian Inference of Pareto Models Based on Information Geometry. *Entropy* **2021**, *23*, 45. [CrossRef] [PubMed]
14. Rohatgi, V.K.; Saleh, A.K.M.E. Estimation of the Common Scale Parameter of Two Pareto Distributions in Censored Samples. *Nav. Res. Logist.* **1987**, *34*, 235–238. [CrossRef]
15. Ghosh, M.; Razmpour, A. Estimation of the Common Location Parameter of Several Exponentials. *Sankhyā Indian J. Stat. Ser. A (1961–2002)* **1984**, *46*, 383–394. Available online: http://xxx.lanl.gov/abs/25050498 (accessed on 4 May 2023).
16. Bartlett, M.S. Approximate Confidence Intervals. *Biometrika* **1953**, *40*, 12–19.
17. Takeuchi, J.; Amari, S. Alpha-Parallel Prior and Its Properties. *IEEE Trans. Inf. Theory* **2005**, *51*, 1011–1023. [CrossRef]
18. Robert, C.P. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed.; Springer Texts in Statistics; Springer: New York, NY, USA, 2007.
19. Jeffreys, H. *Theory of Probability*, 3rd ed.; Clarendon Press: Oxford, UK, 1961.
20. Nomizu, K.; Sasaki, T. Affine Differential Geometry. In *Cambridge Tracts in Mathematics*; Cambridge University Press: Cambridge, UK, 1994; Volume 111.
21. Arnold, B.C. *Pareto Distributions*. In *Statistical Distributions in Scientific Work*; International Co-operative Publishing House: Burtonsville, MD, USA, 1983; Volume 5.

22. Akahira, M.; Ohyauchi, N. Second-Order Asymptotic Loss of the MLE of a Truncation Parameter for a Truncated Exponential Family of Distributions. *Commun. Stat. Theory Methods* **2017**, *46*, 6085–6097. [CrossRef]
23. Akahira, M. Second Order Asymptotic Variance of the Bayes Estimator of a Truncation Parameter for a One-Sided Truncated Exponential Family of Distributions. *J. Jpn. Stat. Soc. (Nihon Tôkei Gakkai Kaihô)* **2016**, *46*, 81–98. [CrossRef]
24. Bernardo, J.M. Reference Posterior Distributions for Bayesian Inference. *J. R. Stat. Soc. Ser. B (Methodol.)* **1979**, *41*, 113–128. [CrossRef]
25. Ghosh, J.K.; Mukerjee, R. Non-Informative Priors. In *Bayesian Statistics, 4 (Peñíscola, 1991)*; Oxford University Press: New York, NY, USA, 1992; pp. 195–210.
26. Ghosal, S. Reference Priors in Multiparameter Nonregular Cases. *Test* **1997**, *6*, 159–186. [CrossRef]