Asymptotic expansion of the risk of maximum likelihood estimator with respect to α-divergence

Yo Sheena[*]

Jan. 2017

For a given parametric probability model, we consider the risk of the maximum likelihood estimator with respect to α-divergence, which includes the special cases of Kullback–Leibler divergence, the Hellinger distance and essentially χ^2-divergence. The asymptotic expansion of the risk is given with respect to sample sizes up to order n^{-2}. Each term in the expansion is expressed with the geometrical properties of the Riemannian manifold formed by the parametric probability model.

1 Introduction

For parametric models of probability distributions, we are naturally concerned with the following comparisons:
1. Comparison of the risks of estimation among different parameters within a same model.
2. Overall comparison of the risks between different models.

To carry out these comparisons, particularly the second, we need some way of measuring the risk of estimation that is common to all parametric probability models. The maximum likelihood method, in which the maximum likelihood estimator (m.l.e.) is plugged into the unknown parameters, is the most common approach, and is applicable

---

[*]Faculty of Economics and Law, Shinshu University; Faculty of Data Science, Shiga University; Visiting Professor of the Institute of Statistical Mathematics.

to any parametric model. In this paper, we choose the m.l.e., and use its risk with respect to a certain loss function.

For comparisons above, particularly the first, the loss function should be independent of the choice of the parameter (coordinate). Consider the binomial distribution model $B(n, p)$, where n is known and we wish to estimate p. Take the quadratic loss function as an example of a parameter-dependent loss function:

$$L(\hat{p}, p) = (\hat{p} - p)^2.$$

The m.l.e. $\hat{p}$ is the sample ratio. The risk of the m.l.e. with respect to this quadratic loss function is given by

$$E_p[(\hat{p} - p)^2] = V(\hat{p}) = p(1 - p)/n,$$

which says that the estimation of this model reaches the highest risk point when $p = 1/2$. In contrast, if we use p^{-1} as a parameter for the model, the risk of the m.l.e. is given by

$$E_{p^{-1}}[(\hat{p}^{-1} - p^{-1})^2] = p^{-3}(1 - p)/n + o(n^{-1}).$$

For a large sample size, the highest risk point is $p = 0$.

Another consideration for the loss function is the invariance with respect to the transformation between the random variables. If $Y(X)$ is a sufficient statistic for the parametric model of a random object X, then the risk of estimation should be measured independently of the choice of Y and X. In particular, when $X \leftrightarrow Y$ is a one-to-one transformation, the parametric models for the distribution of X and Y are essentially equivalent. We wish to measure the risk of estimating the model without being concerned by the form in which the observations were acquired.

Taking these considerations into account, *f-divergence* is a natural loss function, as it satisfies both parameter independence and transformation invariance. It also has other favorable properties (see, e.g., Chapter 9 of Vajda [21]), and has been widely used in engineering problems (e.g. [13], [16]). Because f-divergence is quite a general class of divergence, we need more specific forms for a concrete result. Here, we focus on *α-divergence*. This is a subclass of f-divergence, but is still a general class of divergence, and includes the well-known Kullback–Leibler divergence ($\alpha = -1$), the Hellinger distance ($\alpha = 0$). It is also equivalent, when $\alpha = -3$, to χ^2-divergence in a small neighborhood of the point where the two probabilities coincide. More importantly, from the perspective of information geometry, α-divergence gives rise to a "dually flat" structure for the manifold of the given parametric model (see Eguchi [8], Amari [3], and Amari and Cichocki [4]).

Consequently, in this paper, we consider the risk of the m.l.e. with respect to α-divergence in estimating the parameter. However, an exact calculation of the risk of the m.l.e is often quite difficult, even if we choose a mathematically easy-to-handle quantity such as Kullback–Leibler divergence. Hence, it would be useful if we had an asymptotic expansion of the risk with respect to the sample size n. The main concern of this paper is to give this expansion for a general α up to order n^{-2}. The result is expressed with the geometrical properties of the Riemannian manifold formed by the

parametric model. The geometric expression of the risk expansion provides an insight into which geometrical property of the model affects the estimation risk. As examples, we calculate the asymptotic risk expansion for some specific families of distributions such as an exponential family and a mixture family.

The most relevant work to this paper is that of Komaki [11] and Corcuera and Giummolè [7]. Both of these papers gave asymptotic expansions of the risk of *predictive distributions* with respect to Kullback–Leibler divergence (Theorem 1 of [11]) and α-divergence (Theorem 3.1 of [7]). They considered a curved exponential family as the presupposed model, and developed a general method for the asymptotic improvement of an *estimative distribution* (that is, a distribution within the model gained by parameter estimation) using a predictive distribution that belongs to a full exponential family but lies outside of the model. Note that the whole class of predictive distributions includes the estimative distributions as special cases, and the distribution with the estimated parameter by the m.l.e. is a typical estimative distribution. Hence, [11] and [7] treat a more general class of estimation procedure than m.l.e. under the framework of curved exponential families. This paper differs in two points from these papers. First, we do not presuppose the exponential families. We gain the main result, Theorem 1 in Section 2, for a general parametric family. Second, we present all the results through geometrical terms, which enables us to understand how the geometrical structure of the model is related to the risk of model estimation.

We now state the formal framework of the problem. First, we consider a parametric family of probability distributions on a space $\mathfrak{X}$ (say, $\mathcal{P}$), which is given by a family of positive-valued densities $f(x; \theta)$ on $\mathfrak{X}$ with respect to a measure μ:

$$\mathcal{P} = \{f(x; \theta) \mid \theta = (\theta^1, \dots, \theta^p), \theta \in \Theta\}, \tag{1}$$

where Θ is an open set in R^p, and $f(x; \theta_1) = f(x; \theta_2)$ almost everywhere if and only if $\theta_1 = \theta_2$. We will treat $\mathcal{P}$ as a Riemannian manifold, and define several geometrical properties on it.

α-divergence $(-\infty < \alpha < \infty)$ between $f(x; \theta_1)$ and $f(x; \theta_2)$ is defined as

$$\overset{\alpha}{D}[\theta_1 : \theta_2] = \begin{cases} \frac{4}{1-\alpha^2}\left\{1 - \int_{\mathfrak{X}} f^{(1-\alpha)/2}(x; \theta_1) f^{(1+\alpha)/2}(x; \theta_2) d\mu\right\}, & \text{if } \alpha \neq \pm 1, \\ \int_{\mathfrak{X}} f(x; \theta_2) \log\left(f(x; \theta_2)/f(x; \theta_1)\right) d\mu, & \text{if } \alpha = 1, \\ \int_{\mathfrak{X}} f(x; \theta_1) \log\left(f(x; \theta_1)/f(x; \theta_2)\right) d\mu, & \text{if } \alpha = -1. \end{cases} \tag{2}$$

As a general divergence property, this satisfies $\overset{\alpha}{D}[\theta_1 : \theta_2] \geq 0$, where the equality holds if and only if $\theta_1 = \theta_2$. It is not symmetric between θ_1 and θ_2. Actually, the following relation holds:

$$\overset{\alpha}{D}[\theta_1 : \theta_2] = \overset{-\alpha}{D}[\theta_2 : \theta_1].$$

If we adopt α-divergence as a loss function between the m.l.e. $\hat{\theta}$ and the true parameter θ, the risk of the m.l.e. is given by

$$\overset{\alpha}{ED}(\theta) \triangleq E_\theta\big[\overset{\alpha}{D}[\hat{\theta}(\boldsymbol{X}) : \theta]\big],$$

3

where $\boldsymbol{X} = (X_1, \ldots, X_n)$ are n independent random samples from the distribution given by $f(x; \theta)$.

The remainder of this paper is organized as follows. In Section 2, we give the main result (Theorem 1) and its corollaries for some distribution families and specific values of α. In Section 3, we give examples of the adaptation of the main result to some specific distribution. The minimum knowledge of the information geometry and some preliminary results used for the derivation for Theorem 1 are presented in the appendix. The lengthy calculation of geometrical properties and the proof of some results are omitted in this paper. For the detailed description of these parts, refer to Sheena [17].

Before closing this section, we state some technical conditions and notation used throughout this paper. We assume that $f(x; \theta)$ is differentiable at least five times with respect to θ, and that differentiation and integration on $\mathfrak{X}$ is always exchangeable. Every expectation that appears in this context is assumed to be finite. (We refer to these conditions as "C.1".) The following notation is used:

$$\partial_i \triangleq \frac{\partial}{\partial \theta^i}$$

$$f_i \triangleq f_i(x; \theta) \triangleq \partial_i f(x; \theta), \quad f_{ij} \triangleq f_{ij}(x; \theta) \triangleq \partial_i \partial_j f(x; \theta), \quad \cdots$$

$$l_i \triangleq l_i(x; \theta) \triangleq \partial_i \log f(x; \theta), \quad l_{ij} \triangleq l_{ij}(x; \theta) \triangleq \partial_i \partial_j \log f(x; \theta), \quad \cdots$$

$$E_\theta[h(x; \theta)] \triangleq \int_{\mathfrak{X}} h(x; \theta) f(x; \theta) d\mu,$$

$$\sum_i \triangleq \sum_{1 \leq i \leq p}, \quad \sum_{i,j} \triangleq \sum_{1 \leq i,j \leq p}, \quad \cdots$$

2 Asymptotic Expansion of the Risk of M.L.E.

In this section we consider the asymptotic expansion of the risk of the maximum likelihood estimator (m.l.e.) with respect to α-divergence.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be n independent random sample from the distribution given by $f(x; \theta)$. The m.l.e. of θ based on X is denoted by $\hat{\theta}(X)$. The expected divergence at θ between $f(x; \hat{\theta}(\boldsymbol{X}))$ and $f(x; \theta)$ is given by

$$\overset{\alpha}{ED}(\theta) \triangleq E_\theta\big[\overset{\alpha}{D}[\hat{\theta}(\boldsymbol{X}) : \theta]\big].$$

Using the expansion of divergence (85) in Section 4.3 , then we have

$$\overset{\alpha}{ED}(\theta)$$

$$= \sum_{i=1}^{p} \big(\epsilon_i \overset{\alpha}{D}[\theta : \theta]\big) E_\theta[(\hat{\theta}^i - \theta^i)] + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \big(\epsilon_i \epsilon_j \overset{\alpha}{D}[\theta : \theta]\big) E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)]$$

$$+ \frac{1}{6} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \big(\epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta : \theta]\big) E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k)]$$

$$+ \frac{1}{24} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} \left(\epsilon_i \epsilon_j \epsilon_k \epsilon_l \overset{\alpha}{D}[\theta:\theta] \right) E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k)(\hat{\theta}^l - \theta^l)]$$

$$+ E_\theta[O(||\hat{\theta} - \theta||^5)]. \tag{3}$$

As we see in Section 4.3, the terms

$$\epsilon_i \overset{\alpha}{D}[\theta:\theta], \ \ \epsilon_i \epsilon_j \overset{\alpha}{D}[\theta:\theta], \ \ \epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta:\theta], \ \ \epsilon_i \epsilon_j \epsilon_k \epsilon_l \overset{\alpha}{D}[\theta:\theta]$$

could be geometrically interpreted. In this section, we will show that the terms

$$E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)], \quad E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k)],$$
$$E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k)(\hat{\theta}^l - \theta^l)] \tag{4}$$

are also interpretable with the geometrical properties of $\mathcal{P}$.

First we evaluate $\bar{\theta}^i \triangleq \hat{\theta}^i - \theta^i$ following the way of Eguchi and Yanagimoto [10]. Let

$$\bar{e}_i(\boldsymbol{X};\theta) \triangleq \frac{1}{n} \sum_{a=1}^{n} \frac{\partial}{\partial \theta^i} \log f(X_a;\theta), \qquad \bar{e}^i(\boldsymbol{X};\theta) \triangleq \sum_{j=1}^{p} g^{ij} \bar{e}_j(\boldsymbol{X};\theta)$$

for $1 \le i \le p$. Since m.l.e. $\hat{\theta}$ maximizes log-likelihood $\sum_{a=1}^{n} \log f(x_a;\theta)$

$$\bar{e}_i(\boldsymbol{X};\hat{\theta}) = 0. \tag{5}$$

Taylor expansion of $\bar{e}_i(\boldsymbol{X};\hat{\theta})$ around θ is given by

$$\bar{e}_i(\boldsymbol{X};\hat{\theta})$$
$$= \bar{e}_i(\boldsymbol{X};\theta) + \sum_j \frac{\partial \bar{e}_i(\boldsymbol{X};\theta)}{\partial \theta^j} \bar{\theta}^j + \frac{1}{2!} \sum_{j,k} \frac{\partial^2 \bar{e}_i(\boldsymbol{X};\theta)}{\partial \theta^j \partial \theta^k} \bar{\theta}^j \bar{\theta}^k$$
$$+ \frac{1}{3!} \sum_{j,k,l} \frac{\partial^3 \bar{e}_i(\boldsymbol{X};\theta)}{\partial \theta^j \partial \theta^k \partial \theta^l} \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l + \frac{1}{4!} \sum_{j,k,l,m} \frac{\partial^4 \bar{e}_i(\boldsymbol{X};\theta_i^*)}{\partial \theta^j \partial \theta^k \partial \theta^l \partial \theta^m} \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l \bar{\theta}^m,$$

where θ_i^* is on the segment $\overline{\theta \hat{\theta}}$. If we add $\sum_j g_{ij}(\theta) \bar{\theta}^j$ on the both sides of the above expansion and use (5), then we have

$$\sum_j g_{ij}(\theta) \bar{\theta}^j$$
$$= \bar{e}_i(\boldsymbol{X};\theta) + \sum_j \left(\frac{\partial \bar{e}_i(\boldsymbol{X};\theta)}{\partial \theta^j} + g_{ij}(\theta) \right) \bar{\theta}^j + \frac{1}{2!} \sum_{j,k} \frac{\partial^2 \bar{e}_i(\boldsymbol{X};\theta)}{\partial \theta^j \partial \theta^k} \bar{\theta}^j \bar{\theta}^k$$
$$+ \frac{1}{3!} \sum_{j,k,l} \frac{\partial^3 \bar{e}_i(\boldsymbol{X};\theta)}{\partial \theta^j \partial \theta^k \partial \theta^l} \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l + \frac{1}{4!} \sum_{j,k,l,m} \frac{\partial^4 \bar{e}_i(\boldsymbol{X};\theta_i^*)}{\partial \theta^j \partial \theta^k \partial \theta^l \partial \theta^m} \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l \bar{\theta}^m.$$

Furthermore if we multiply the both sides with g^{is} and sum them up over i from 1 to p, then we have

$$\bar{\theta}^s = \bar{e}^s + \sum_j A_j^s \bar{\theta}^j + \sum_{j,k} B_{jk}^s \bar{\theta}^j \bar{\theta}^k + \sum_{j,k} \bar{B}_{jk}^s \bar{\theta}^j \bar{\theta}^k + \sum_{j,k,l} C_{jkl}^s \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l + \sum_{jkl} \bar{C}_{jkl}^s \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l$$
$$+ \sum_{j,k,l,m} D_{jklm}^s \bar{\theta}^j \bar{\theta}^k \bar{\theta}^l \bar{\theta}^m, \tag{6}$$

where we used the following notations: For $1 \leq j, k, l, m, s \leq p$,

$$A_j^s(\boldsymbol{X}; \theta) \triangleq \sum_{i=1}^p g^{is}(\theta) \left(\frac{\partial \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j} + g_{ij}(\theta) \right)$$

$$B_{jk}^s(\boldsymbol{X}; \theta) \triangleq \frac{1}{2} \sum_{i=1}^p g^{is}(\theta) \left(\frac{\partial^2 \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j \partial \theta^k} - E_\theta \left[\frac{\partial^2 \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j \partial \theta^k} \right] \right)$$

$$\bar{B}_{jk}^s(\theta) \triangleq \frac{1}{2} \sum_{i=1}^p g^{is}(\theta) E_\theta \left[\frac{\partial^2 \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j \partial \theta^k} \right]$$

$$\bar{C}_{jkl}^s(\theta) \triangleq \frac{1}{3!} \sum_{i=1}^p g^{is}(\theta) E_\theta \left[\frac{\partial^3 \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j \partial \theta^k \partial \theta^l} \right]$$

$$C_{jkl}^s(\boldsymbol{X}; \theta) \triangleq \frac{1}{3!} \sum_i g^{is}(\theta) \left(\frac{\partial^3 \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j \partial \theta^k \partial \theta^l} - E_\theta \left[\frac{\partial^3 \bar{e}_i(\boldsymbol{X}; \theta)}{\partial \theta^j \partial \theta^k \partial \theta^l} \right] \right),$$

$$D_{jklm}^s(\boldsymbol{X}; \theta) \triangleq \frac{1}{4!} \sum_i g^{is}(\theta) \frac{\partial^4 \bar{e}_i(\boldsymbol{X}; \theta_i^*)}{\partial \theta^j \partial \theta^k \partial \theta^l \partial \theta^m}.$$

Here we impose the moment conditions as follows. The suitably higher-order joint moments composed of the following variables are bounded with respect to n;

$$\sqrt{n}\, \bar{\theta}^s(\boldsymbol{X}; \theta), \quad \sqrt{n}\, \bar{e}^s(\boldsymbol{X}; \theta), \quad \sqrt{n}\, A_j^s(\boldsymbol{X}; \theta), \quad \sqrt{n}\, B_{jk}^s(\boldsymbol{X}; \theta),$$
$$\sqrt{n}\, C_{jkl}^s(\boldsymbol{X}; \theta), \quad D_{jklm}^s(\boldsymbol{X}; \theta), \tag{7}$$

where $1 \leq j, k, l, m, s \leq p$. (We refer to this condition as "C.2".) We obtain the following results (for the proof, see [17]). For $1 \leq s \leq p$,

$$\bar{\theta}^s = \bar{e}^s + \sum_j A_j^s \bar{e}^j + \sum_{j,k} \bar{B}_{jk}^s \bar{e}^j \bar{e}^k + \sum_{i,j} A_j^s A_i^j \bar{e}^i + \sum_{i,j,k} A_j^s \bar{B}_{ik}^j \bar{e}^i \bar{e}^k$$
$$+ \sum_{j,k} B_{jk}^s \bar{e}^j \bar{e}^k + 2 \sum_{i,j,k} \bar{B}_{jk}^s A_i^k \bar{e}^i \bar{e}^j + 2 \sum_{i,j,k,l} \bar{B}_{jk}^s \bar{B}_{il}^k \bar{e}^i \bar{e}^j \bar{e}^l \tag{8}$$
$$+ \sum_{j,k,l} \bar{C}_{jkl}^s \bar{e}^j \bar{e}^k \bar{e}^l + Re(4),$$

where $Re(4)$ is the polynomial with respect to the variables $\bar{\theta}^s$, $\bar{e}^s$, A_j^s, B_{jk}^s, C_{jkl}^s, D_{jklm}^s $(1 \leq j, k, l, m, s \leq p)$, and each term is of at least fourth order with respect to $\bar{\theta}^s$, $\bar{e}^s$, A_j^s, B_{jk}^s, C_{jkl}^s $(1 \leq j, k, l, s \leq p)$.

Combining this evaluation with Lemma 2 in Section 4.2, we can express three expectations in (4) with geometrical terms (for the detailed calculation, see [17]). The results are given as follows. For achieving notational brevity, we use Einstein summation convention (the summation is carried out as every pair of upper and lower index moves

from 1 to p).

$$E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)]$$
$$= n^{-1}g^{ij} + n^{-2}\times$$
$$\Bigg[g^{sj}g^{it}g^{lm}\langle \overset{e}{A}_{sl}, \overset{m}{A}_{tm} - \overset{e}{A}_{tm}\rangle + g^{si}g^{jt}g^{lm}\langle \overset{e}{A}_{sl}, \overset{m}{A}_{tm} - \overset{e}{A}_{tm}\rangle$$

$$+ g^{sj}g^{it}g^{lm}\overset{e}{\Gamma}_{sl,u}\big(\overset{m}{\Gamma}{}^u_{tm} - \overset{e}{\Gamma}{}^u_{tm}\big) + g^{si}g^{jt}g^{lm}\overset{e}{\Gamma}_{sl,u}\big(\overset{m}{\Gamma}{}^u_{tm} - \overset{e}{\Gamma}{}^u_{tm}\big)$$

$$+ \bar{B}^j_{lm}g^{ik}g^{ls}g^{mt}\big(\overset{m}{\Gamma}_{ks,t} - \overset{e}{\Gamma}_{ks,t}\big) + \bar{B}^i_{lm}g^{jk}g^{ls}g^{mt}\big(\overset{m}{\Gamma}_{ks,t} - \overset{e}{\Gamma}_{ks,t}\big)$$

$$+ g^{jk}g^{lu}g^{is}g^{mt}\big(\langle \overset{e}{A}_{kl}, \overset{e}{A}_{um}\rangle g_{st} + \overset{e}{\Gamma}{}^v_{kl}\overset{e}{\Gamma}{}^w_{um}g_{vw}g_{st} + \overset{e}{\Gamma}_{kl,s}\overset{e}{\Gamma}_{um,t} + \overset{e}{\Gamma}_{kl,t}\overset{e}{\Gamma}_{um,s}\big)$$

$$+ g^{ik}g^{lu}g^{js}g^{mt}\big(\langle \overset{e}{A}_{kl}, \overset{e}{A}_{um}\rangle g_{st} + \overset{e}{\Gamma}{}^v_{kl}\overset{e}{\Gamma}{}^w_{um}g_{vw}g_{st} + \overset{e}{\Gamma}_{kl,s}\overset{e}{\Gamma}_{um,t} + \overset{e}{\Gamma}_{kl,t}\overset{e}{\Gamma}_{um,s}\big)$$

$$+ \bar{B}^l_{ms}g^{jk}g^{it}g^{mu}g^{sv}\big(\overset{e}{\Gamma}_{kl,t}g_{uv} + \overset{e}{\Gamma}_{kl,u}g_{tv} + \overset{e}{\Gamma}_{kl,v}g_{tu}\big)$$

$$+ \bar{B}^l_{ms}g^{ik}g^{jt}g^{mu}g^{sv}\big(\overset{e}{\Gamma}_{kl,t}g_{uv} + \overset{e}{\Gamma}_{kl,u}g_{tv} + \overset{e}{\Gamma}_{kl,v}g_{tu}\big)$$

$$+ \frac{1}{2}g^{js}g^{it}g^{lu}g^{mv}\big\{\big((\partial_m\overset{e}{\Gamma}{}^k_{sl})g_{tk} + \overset{e}{\Gamma}{}^k_{sl}\overset{e}{\Gamma}_{km,t} - \langle \overset{e}{A}_{sl}, \overset{m}{A}_{mt}\rangle\big)g_{uv}$$

$$+ \big((\partial_m\overset{e}{\Gamma}{}^k_{sl})g_{uk} + \overset{e}{\Gamma}{}^k_{sl}\overset{e}{\Gamma}_{km,u} - \langle \overset{e}{A}_{sl}, \overset{m}{A}_{mu}\rangle\big)g_{tv}$$

$$+ \big((\partial_m\overset{e}{\Gamma}{}^k_{sl})g_{vk} + \overset{e}{\Gamma}{}^k_{sl}\overset{e}{\Gamma}_{km,v} - \langle \overset{e}{A}_{sl}, \overset{m}{A}_{mv}\rangle\big)g_{ut}\big\}$$

$$+ \frac{1}{2}g^{is}g^{jt}g^{lu}g^{mv}\big\{\big((\partial_m\overset{e}{\Gamma}{}^k_{sl})g_{tk} + \overset{e}{\Gamma}{}^k_{sl}\overset{e}{\Gamma}_{km,t} - \langle \overset{e}{A}_{sl}, \overset{m}{A}_{mt}\rangle\big)g_{uv}$$

$$+ \big((\partial_m\overset{e}{\Gamma}{}^k_{sl})g_{uk} + \overset{e}{\Gamma}{}^k_{sl}\overset{e}{\Gamma}_{km,u} - \langle \overset{e}{A}_{sl}, \overset{m}{A}_{mu}\rangle\big)g_{tv}$$

$$+ \big((\partial_m\overset{e}{\Gamma}{}^k_{sl})g_{vk} + \overset{e}{\Gamma}{}^k_{sl}\overset{e}{\Gamma}_{km,v} - \langle \overset{e}{A}_{sl}, \overset{m}{A}_{mv}\rangle\big)g_{ut}\big\}$$

$$+ 2\bar{B}^j_{lm}g^{mt}g^{iu}g^{lv}g^{sw}\big(\overset{e}{\Gamma}_{st,u}g_{vw} + \overset{e}{\Gamma}_{st,v}g_{uw} + \overset{e}{\Gamma}_{st,w}g_{uv}\big)$$

$$+ 2\bar{B}^i_{lm}g^{mt}g^{ju}g^{lv}g^{sw}\big(\overset{e}{\Gamma}_{st,u}g_{vw} + \overset{e}{\Gamma}_{st,v}g_{uw} + \overset{e}{\Gamma}_{st,w}g_{uv}\big)$$

$$+ 2\bar{B}^j_{lm}\bar{B}^m_{st}g^{ik}g^{lu}g^{sv}g^{tw}\big(g_{ku}g_{vw} + g_{kv}g_{uw} + g_{kw}g_{uv}\big)$$

$$+ 2\bar{B}^i_{lm}\bar{B}^m_{st}g^{jk}g^{lu}g^{sv}g^{tw}\big(g_{ku}g_{vw} + g_{kv}g_{uw} + g_{kw}g_{uv}\big)$$

$$+ \bar{C}^j_{lmt}g^{ik}g^{ls}g^{mu}g^{tv}\big(g_{ks}g_{uv} + g_{ku}g_{sv} + g_{kv}g_{su}\big)$$

$$+ \bar{C}^i_{lmt}g^{jk}g^{ls}g^{mu}g^{tv}\big(g_{ks}g_{uv} + g_{ku}g_{sv} + g_{kv}g_{su}\big)$$

$$+ g^{ik}g^{js}g^{lt}g^{mu}\big(\langle \overset{e}{A}_{kl}, \overset{e}{A}_{sm}\rangle g_{tu} + \overset{e}{\Gamma}{}^v_{kl}\overset{e}{\Gamma}{}^w_{sm}g_{vw}g_{tu} + \overset{e}{\Gamma}_{kl,t}\overset{e}{\Gamma}_{sm,u} + \overset{e}{\Gamma}_{kl,u}\overset{e}{\Gamma}_{sm,t}\big)$$

$$+ \bar{B}^j_{st}g^{ik}g^{lu}g^{sv}g^{tw}\big(\overset{e}{\Gamma}_{kl,u}g_{vw} + \overset{e}{\Gamma}_{kl,v}g_{uw} + \overset{e}{\Gamma}_{kl,w}g_{uv}\big)$$

$$+ \bar{B}^i_{st}g^{jk}g^{lu}g^{sv}g^{tw}\big(\overset{e}{\Gamma}_{kl,u}g_{vw} + \overset{e}{\Gamma}_{kl,v}g_{uw} + \overset{e}{\Gamma}_{kl,w}g_{uv}\big)$$

$$+ \bar{B}^i_{lm}\bar{B}^j_{st}g^{lk}g^{mu}g^{sv}g^{tw}\big(g_{ku}g_{vw} + g_{kv}g_{uw} + g_{kw}g_{uv}\big)\Bigg] + O(n^{-5/2}). \tag{9}$$

$$E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k)]$$

$$= n^{-2}\big\{ g^{is}g^{jt}g^{ku}\big(\overset{m}{\Gamma}_{st,u} - \overset{e}{\Gamma}_{st,u}\big)$$

$$+ \overset{e}{\Gamma}{}^{s}_{st} g^{it}g^{jk} + \overset{e}{\Gamma}{}^{j}_{st} g^{sk}g^{it} + \overset{e}{\Gamma}{}^{k}_{st} g^{sj}g^{it}$$

$$+ \overset{e}{\Gamma}{}^{s}_{st} g^{jt}g^{ik} + \overset{e}{\Gamma}{}^{i}_{st} g^{sk}g^{jt} + \overset{e}{\Gamma}{}^{k}_{st} g^{si}g^{jt}$$

$$+ \overset{e}{\Gamma}{}^{s}_{st} g^{kt}g^{ji} + \overset{e}{\Gamma}{}^{j}_{st} g^{si}g^{kt} + \overset{e}{\Gamma}{}^{i}_{st} g^{sj}g^{kt}$$

$$+ \bar{B}^{i}_{st}\big(g^{st}g^{jk} + g^{js}g^{kt} + g^{ks}g^{jt}\big)$$

$$+ \bar{B}^{j}_{st}\big(g^{st}g^{ik} + g^{is}g^{kt} + g^{ks}g^{it}\big)$$

$$+ \bar{B}^{k}_{st}\big(g^{st}g^{ji} + g^{js}g^{it} + g^{is}g^{jt}\big)\big\} + O(n^{-5/2}). \tag{10}$$

$$E_\theta[(\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k)(\hat{\theta}^l - \theta^l)] = n^{-2}\big(g^{ij}g^{kl} + g^{ik}g^{jl} + g^{il}g^{jk}\big) + O(n^{-5/2}). \tag{11}$$

From (86)–(88), (93) in Section 4.3 and (9)–(11), we obtain the following result (for the detailed calculation, see [17]).

Theorem 1. *Under the conditions C.1 and C.2, the following expansion holds.*

$$\overset{\alpha}{ED}$$

$$= \frac{p}{2n} + \frac{1}{24n^2}$$

$$\times \Big[(\alpha')^2 \big\{ 3\overset{e}{F} + 3T^{ijk}T_{ijk} - 6\langle \overset{e}{A}{}^{j}_{i}, (\overset{m}{A}{}^{i}_{j} - \overset{e}{A}{}^{i}_{j})\rangle - 3\langle \overset{e}{A}{}^{i}_{i}, (\overset{m}{A}{}^{j}_{j} - \overset{e}{A}{}^{j}_{j})\rangle + 3p^2 + 6p \big\}$$

$$+ \alpha' \big\{ 3\overset{e}{F} - 5T^{ijk}T_{ijk} - 6T^{i}_{is}T^{js}_{j} + 6\langle \overset{e}{A}{}^{j}_{i}, (\overset{m}{A}{}^{i}_{j} - \overset{e}{A}{}^{i}_{j})\rangle + 3\langle \overset{e}{A}{}^{i}_{i}, (\overset{m}{A}{}^{j}_{j} - \overset{e}{A}{}^{j}_{j})\rangle$$

$$- 3p^2 - 6p \big\}$$

$$+ 12\langle \overset{e}{A}{}^{i}_{j}, \overset{e}{A}{}^{j}_{i}\rangle - 2\langle \overset{e}{A}{}^{i}_{j}, \overset{m}{A}{}^{j}_{i}\rangle - \langle \overset{e}{A}{}^{i}_{i}, \overset{m}{A}{}^{j}_{j}\rangle + T_{ijk}T^{ijk} + 9T^{i}_{is}T^{js}_{j} + 8\overset{e}{R}_{ij}{}^{ij} - 9\overset{e}{F} \Big]$$

$$+ o(n^{-2}), \tag{12}$$

where $\alpha' = (1 - \alpha)/2$, and for $1 \le i, j, k, l \le p$,

$$\big(g^{ij}(\theta)\big) \triangleq \big(g_{ij}(\theta)\big)^{-1}, \tag{13}$$

$$T_{ijk}(\theta) \triangleq \overset{m}{\Gamma}_{ij,k}(\theta) - \overset{e}{\Gamma}_{ij,k}(\theta) \tag{14}$$

$$T^{k}_{ij}(\theta) \triangleq \sum_{j} T_{ijl}(\theta)g^{lk}(\theta), \tag{15}$$

$$T^{jk}_{i}(\theta) \triangleq \sum_{l,m} T_{iml}(\theta)g^{mj}(\theta)g^{lk}(\theta), \tag{16}$$

$$T^{ijk}(\theta) \triangleq \sum_{t,m,l} T_{tml}(\theta)g^{ti}(\theta)g^{mj}(\theta)g^{lk}(\theta), \tag{17}$$

$$\overset{e}{R}_{ij}{}^{kl}(\theta) \triangleq \sum_{s} \overset{e}{R}_{ijs}{}^{l}(\theta)g^{sk}(\theta), \tag{18}$$

9

$$\overset{e}{A}{}_i^j \triangleq \sum_k \overset{e}{A}{}_{ik}\, g^{kj}, \qquad \overset{m}{A}{}_i^j \triangleq \sum_k \overset{m}{A}{}_{ik}\, g^{kj}, \tag{19}$$

$$\overset{\alpha}{F}(\theta) \triangleq \sum_{i,k,s} g^{ks}(\theta)\partial_s T_{ik}^i(\theta) - \sum_{i,j,s,t} g^{ti}(\theta)\overset{\alpha}{\Gamma}_{it,s}(\theta)T_j^{js}(\theta), \quad \overset{e}{F}(\theta) \triangleq \overset{1}{F}(\theta), \tag{20}$$

which are the variations from the following fundamental geometric properties of Riemannian manifold $\mathcal{P}$ (for their formal definitions, see Section 4.1);

$g_{ij}(\theta)$: *Fisher information metric,*

$\overset{e}{\Gamma}_{ij,k}(\theta),\ \overset{m}{\Gamma}_{ij,k}(\theta)$: *Christoffel's second symbol for e-connection and m-connection,*

$\overset{e}{R}_{ijk}{}^l(\theta)$: *Riemann curvature for e-connection*

$\overset{e}{A}_{ij}(\theta),\ \overset{m}{A}_{ij}(\theta)$: *Second fundamental form for e-connection and m-connection.*

Roughly speaking, $\overset{e}{\Gamma}_{ij,k}$, $\overset{m}{\Gamma}_{ij,k}$ and $\overset{e}{R}_{ijk}{}^l$ give us information on the *intrinsic curvature* of $\mathcal{P}$, while $\overset{e}{A}_{ij}$, $\overset{m}{A}_{ij}$ tell us how the manifold $\mathcal{P}$ is located in the ambient space (*extrinsic curvature*). Since $\mathcal{P}$ is torsion-free, the following equivalence holds;

$\mathcal{P}$ is intrinsically e-flat $\iff \mathcal{P}$ is intrinsically m-flat

$$\iff \overset{e}{\Gamma}_{ij,k}(\theta) = 0 \text{ for some coordinate system}$$

$$\iff \overset{m}{\Gamma}_{ij,k}(\theta) = 0 \text{ for some coordinate system}$$

$$\iff \overset{e}{R}_{ijk}{}^l(\theta) = 0 \text{ for any(some) coordinate system.}$$

If $\overset{e}{A}_{ij}(\theta) = 0$ $\left(\overset{m}{A}_{ij}(\theta) = 0\right)$ for any(some) coordinate system, it means $\mathcal{P}$ is extrinsically e-flat (m-flat).

All properties from (13) to (19) are tensors. $\overset{e}{F}$ is parameter invariant (see [17] for the proof). Consequently every term in the right-hand side of (12) is parameter invariant as is expected from the parameter independence of α-divergence. This means that we can choose any coordinate system with which we can easily calculate the terms in (12). For example, if we have another coordinate system $\eta \triangleq (\eta_\alpha, \eta_\beta, \eta_\gamma, \ldots)$ for $\mathcal{P}$, we can choose to calculate such a term as $T_{\alpha\beta\gamma}T^{\alpha\beta\gamma}$ instead of $T_{ijk}T^{ijk}$.

We easily notice that $T_{ijk}T^{ijk}$ and $T_{is}^i T_j^{js}$ is nonnegative (see [17] for the proof), but other terms in the bracket in (12) could be negative. As we will see in Section 3, the n^{-2} term could be negative.

Note that the n^{-1} term equals $p/2n$, hence the risk in estimating a model by m.l.e. is primarily determined by the number of the parameters, in other words, "model complexity". The number of the parameters p also appears in A.I.C. as the penalty to the model complexity. This is natural since A.I.C. (and some other information criteria for model selection) is considered to be an estimator of $\overset{-1}{ED}$.

A geometrical expression (12) immediately leads us to the simplified form for an exponential family or a mixture family. The canonical form of an exponential family is given by

$$f(x;\theta) = \exp\big(\theta^1 h_1(x) + \cdots + \theta^p h_p(x) - \psi(\theta)\big), \tag{21}$$

where $h_i(x)(i = 1, \ldots, p)$ is a measurable function on $\mathfrak{X}$, while the one for a mixture family is given by

$$f(x;\theta) = \theta^1 g_1(x) + \cdots + \theta^p g_p(x) + (1 - \sum_{i=1}^{p}\theta^i)g_0(x), \tag{22}$$

where $g_i(x)(i = 0, \ldots, p)$ is a probability density function. These families are characterized respectively as being extrinsically e-flat and m-flat. Namely $\overset{e}{a}_{ij}(x;\theta) = 0$ for an exponential family, and $\overset{m}{a}_{ij}(x;\theta) = 0$ for a mixture family. Furthermore an exponential family is intrinsically e and m-flat. A mixture family is also e and m-flat. This means $\overset{e}{R}_{ij}{}^{ij}$ vanishes for both families.

Consequently we have the following corollaries.

Corollary 1. *If the model $\mathcal{P}$ is an exponential family,*

$$\overset{\alpha}{ED}$$

$$= \frac{p}{2n} + \frac{1}{24n^2}$$

$$\times \Big[(\alpha')^2\big\{3\overset{e}{F} + 3T^{ijk}T_{ijk} + 3p^2 + 6p\big\}$$

$$\quad + \alpha'\big\{3\overset{e}{F} - 5T^{ijk}T_{ijk} - 6T_{is}^i T_j^{js} - 3p^2 - 6p\big\}$$

$$\quad + T_{ijk}T^{ijk} + 9T_{is}^i T_j^{js} - 9\overset{e}{F}\Big] + o(n^{-2}). \tag{23}$$

Proof. If the model $\mathcal{P}$ is an exponential family, the terms

$$\langle \overset{e}{A}_i^j, \overset{e}{A}_j^i \rangle, \qquad \langle \overset{e}{A}_i^i, \overset{e}{A}_j^j \rangle, \qquad \langle \overset{e}{A}_i^j, \overset{m}{A}_j^i \rangle, \qquad \langle \overset{e}{A}_i^i, \overset{m}{A}_j^j \rangle, \qquad \overset{e}{R}_{ij}{}^{ij}$$

vanish. □

Corollary 2. *If $\mathcal{P}$ is a mixture family,*

$$\overset{\alpha}{ED}$$

$$= \frac{p}{2n} + \frac{1}{24n^2}$$

$$\times \Big[(\alpha')^2\big\{3\overset{e}{F} + 3T^{ijk}T_{ijk} + 6\langle \overset{e}{A}_j^i, \overset{e}{A}_i^j \rangle + 3\langle \overset{e}{A}_i^i, \overset{e}{A}_j^j \rangle + 3p^2 + 6p\big\}$$

$$\quad + \alpha'\big\{3\overset{e}{F} - 5T^{ijk}T_{ijk} - 6T_{is}^i T_j^{js} - 6\langle \overset{e}{A}_j^i, \overset{e}{A}_i^j \rangle - 3\langle \overset{e}{A}_i^i, \overset{e}{A}_j^j \rangle - 3p^2 - 6p\big\}$$

$$\quad + 12\langle \overset{e}{A}_j^i, \overset{e}{A}_i^j \rangle + T_{ijk}T^{ijk} + 9T_{is}^i T_j^{js} - 9\overset{e}{F}\Big] + o(n^{-2}). \tag{24}$$

11

Proof. If $\mathcal{P}$ is a mixture family, the terms

$$\langle \overset{e}{A}{}_i^j, \overset{m}{A}{}_j^i \rangle, \qquad \langle \overset{e}{A}{}_i^i, \overset{m}{A}{}_j^j \rangle, \qquad \overset{e}{R}{}_{ij}{}^{ij}$$

vanish. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

It is notable that the asymptotic risk for an exponential family depends only on the intrinsic properties of the family.

For a specific α, the following result holds: If $\alpha = -1$, then $\overset{\alpha}{D}[\theta_1 : \theta_2]$ is Kullback-Leibler divergence, and

$$
\begin{aligned}
\overset{-1}{ED} \\
&= \frac{p}{2n} + \frac{1}{24n^2} \\
&\times \left[-3\overset{e}{F} - T^{ijk}T_{ijk} + 3T_{is}^i T_j^{js} + 12\langle \overset{e}{A}{}_j^i, \overset{e}{A}{}_i^j \rangle - 2\langle \overset{e}{A}{}_j^i, \overset{m}{A}{}_i^j \rangle - \langle \overset{e}{A}{}_i^i, \overset{m}{A}{}_j^j \rangle + 8\overset{e}{R}{}_{ij}{}^{ij} \right] \\
&+ o(n^{-2});
\end{aligned}
\tag{25}
$$

if $\alpha = 0$, then $\overset{\alpha}{D}[\theta_1 : \theta_2]$ is equivalent to Hellinger-distance, and

$$
\begin{aligned}
\overset{0}{ED} \\
&= \frac{p}{2n} + \frac{1}{24n^2} \\
&\times \left[-(27/4)\overset{e}{F} - (3/4)T^{ijk}T_{ijk} + 6T_{is}^i T_j^{js} + (21/2)\langle \overset{e}{A}{}_j^i, \overset{e}{A}{}_i^j \rangle - (3/4)\langle \overset{e}{A}{}_i^i, \overset{e}{A}{}_j^j \rangle \right. \\
&\qquad \left. - (1/2)\langle \overset{e}{A}{}_j^i, \overset{m}{A}{}_i^j \rangle - (1/4)\langle \overset{e}{A}{}_i^i, \overset{m}{A}{}_j^j \rangle + 8\overset{e}{R}{}_{ij}{}^{ij} - (3/4)p^2 - (3/2)p \right] \\
&+ o(n^{-2});
\end{aligned}
\tag{26}
$$

if $\alpha = -3$, then $\overset{\alpha}{D}[\theta_1 : \theta_2]$ is asymptotically equivalent to χ^2-divergence, and

$$
\begin{aligned}
\overset{-3}{ED} \\
&= \frac{p}{2n} + \frac{1}{24n^2} \\
&\times \left[9\overset{e}{F} + 3T^{ijk}T_{ijk} - 3T_{is}^i T_j^{js} + 24\langle \overset{e}{A}{}_j^i, \overset{e}{A}{}_i^j \rangle + 6\langle \overset{e}{A}{}_i^i, \overset{e}{A}{}_j^j \rangle - 14\langle \overset{e}{A}{}_j^i, \overset{m}{A}{}_i^j \rangle - 7\langle \overset{e}{A}{}_i^i, \overset{m}{A}{}_j^j \rangle \right. \\
&\qquad \left. + 8\overset{e}{R}{}_{ij}{}^{ij} + 6p^2 + 12p \right] \\
&+ o(n^{-2}).
\end{aligned}
\tag{27}
$$

Generally speaking, the components in the n^{-2} term are not explicitly gained. For the feasibility of the calculation, the expression of these components by the expectation of the derivatives of log-likelihood is useful. Define the following notations; for $1 \leq i, j, k, l \leq p$,

$$L_{(ij)} \triangleq E_\theta[l_{ij}], \quad L_{ij} \triangleq E_\theta[l_i l_j],$$

$$L_{(ij)k} \triangleq E_\theta[l_{ij} l_k], \quad L_{ijk} \triangleq E_\theta[l_i l_j l_k]$$

$$L_{(ij)(kl)} \triangleq E_\theta[l_{ij} l_{kl}], \quad L_{(ijk)l} \triangleq E_\theta[l_{ijk} l_l], \quad L_{(ij)kl} \triangleq E_\theta[l_{ij} l_k l_l], \quad L_{ijkl} \triangleq E_\theta[l_i l_j l_k l_l],$$

$$L11 = g^{ij} g^{kl} L_{(il)jk}, \quad L12 = g^{ij} g^{kl} L_{(ij)kl}, \quad L13 = g^{ij} g^{kl} L_{ijkl},$$

$$L14 = g^{ij} g^{kl} L_{(ik)(jl)}, \quad L15 = g^{ij} g^{kl} L_{(ij)(kl)},$$

$$L21 = g^{ij} g^{kl} g^{su} L_{(ik)s} L_{jlu}, \quad L22 = g^{ij} g^{kl} g^{su} L_{(ij)k} L_{lsu}, \quad L23 = g^{ij} g^{kl} g^{su} L_{iks} L_{jlu},$$

$$L24 = g^{ij} g^{kl} g^{su} L_{ijk} L_{lsu}, \quad L25 = g^{ij} g^{kl} g^{su} L_{(ik)s} L_{(jl)u}, \quad L26 = g^{ij} g^{kl} g^{su} L_{(ij)k} L_{(su)l}.$$

Then we have the following equations (see [17] for the proof).

$$g_{ij} = L_{ij} = -L_{(ij)}, \tag{28}$$

$$\overset{\alpha}{F} = g^{ij} g^{ks} \big(2L_{(is)jk} + L_{(ks)ij} + L_{ijks} \big)$$
$$\quad - g^{ks} g^{uj} g^{li} L_{ijk} \big(2L_{(su)l} + L_{sul} \big)$$
$$\quad - g^{ti} g^{uj} g^{ks} \big(L_{(it)s} + ((1-\alpha)/2) L_{its} \big) L_{juk}$$
$$\quad = 2L11 + L12 + L13 - 2L21 - L23 - L22 - \alpha' L24, \tag{29}$$

$$T_{ijk} T^{ijk} = L_{ijk} L_{stu} g^{is} g^{jt} g^{ku} = L23, \tag{30}$$

$$T_{is}^i T_j^{js} = L_{ijk} L_{stu} g^{ij} g^{st} g^{uk} = L24, \tag{31}$$

$$\overset{e}{R}_{ij}{}^{ij} = g^{ij} g^{sk} \big(L_{(ki)(js)} - L_{(ij)(ks)} + L_{(ki)js} - L_{(ij)ks} \big)$$
$$\quad + g^{sk} g^{ti} g^{uj} \big(-L_{(ki)j} L_{(st)u} + L_{(it)s} L_{(uj)k} + L_{sit} L_{(uj)k} - L_{stu} L_{(ij)k} \big)$$
$$\quad = L14 - L15 + L11 - L12 - L25 + L26 + L22 - L21, \tag{32}$$

$$\langle \overset{e}{A}{}_i^j, \overset{e}{A}{}_j^i \rangle = g^{jk} g^{li} L_{(ik)(jl)} - g^{jk} g^{li} g^{st} L_{(ik)s} L_{(jl)t} - p$$
$$\quad = L14 - L25 - p, \tag{33}$$

$$\langle \overset{e}{A}{}_i^i, \overset{e}{A}{}_j^j \rangle = g^{ik} g^{jl} L_{(ik)(jl)} - g^{ik} g^{jl} g^{st} L_{(ik)s} L_{(jl)t} - p^2$$
$$\quad = L15 - L26 - p^2, \tag{34}$$

$$\langle \overset{e}{A}{}_i^j, \overset{m}{A}{}_j^i \rangle = g^{jk} g^{li} L_{(ik)jl} + g^{jk} g^{li} L_{(ik)(jl)}$$
$$\quad\quad - g^{jk} g^{li} g^{st} L_{(ik)s} L_{(jl)t} - g^{jk} g^{li} g^{st} L_{(ik)s} L_{jlt}$$
$$\quad = L11 + L14 - L25 - L21, \tag{35}$$

$$\langle \overset{e}{A}{}_i^i, \overset{m}{A}{}_j^j \rangle = g^{ik} g^{jl} L_{(ik)jl} + g^{ik} g^{jl} L_{(ik)(jl)}$$
$$\quad\quad - g^{ik} g^{jl} g^{st} L_{(ik)s} L_{(jl)t} - g^{ik} g^{jl} g^{st} L_{(ik)s} L_{jlt}$$
$$\quad = L12 + L15 - L26 - L22. \tag{36}$$

Using these expressions, we could find the value of the components of the n^{-2} term by simulation.

3 Examples

In this section, we take three parametric models as the examples and investigate the concrete form of $\overset{\alpha}{ED}$ up to the n^{-2} term.

−Example 1−

First we consider a discrete model, that is, a multinomial distribution. Consider the family consisting of $p+1$ dimensional multinomial distributions given by

$$P(X = x_i) = m_i, \quad i = 0, 1, \ldots, p,$$

where $\sum_{i=0}^{p} m_i = 1$. We use $m \triangleq (m_1, \ldots, m_p)$ as a free parameter.

The multinomial distribution is an exponential family, then from (23), we notice that we only need to calculate three terms $T_{ijk}T^{ijk}$, $T_{is}^i T_j^{js}$, $\overset{e}{F}$. Because of the following relation (for the proof, see [17]),

$$\overset{e}{F} = \overset{m}{F} + T_{is}^i T_j^{js}, \qquad \overset{m}{F} \triangleq \overset{-1}{F}, \tag{37}$$

we only have to calculate $\overset{m}{F}$ instead of $\overset{e}{F}$. Actually we have the following results (for the proof, see [17]);

$$T_{ijk}T^{ijk} = M - 3p - 1 \tag{38}$$

$$T_{is}^i T_j^{js} = M - (p+1)^2 \tag{39}$$

$$\overset{m}{F} = -M + p + 1, \tag{40}$$

where $M \triangleq \sum_{t=0}^{p} m_t^{-1}$. Consequently we have

$$
\begin{aligned}
\overset{\alpha}{ED}&(m) \\
&= \frac{p}{2n} + \frac{1}{24n^2}\left\{(\alpha')^2(3M - 6p - 3) + \alpha'(-11M + 18p + 11) + 10M - 12p - 10\right\} + o(n^{-2}) \\
&= \frac{p}{2n} + \frac{1}{24n^2}\left\{\frac{(1-\alpha)^2}{4}(3M - 6p - 3) + \frac{1-\alpha}{2}(-11M + 18p + 11) + 10M - 12p - 10\right\} \\
&\quad + o(n^{-2}) \\
&= \frac{p}{2n} + \frac{1}{96n^2}\left\{(3\alpha^2 + 16\alpha + 21)(M - 1) + (-6\alpha^2 - 24\alpha - 18)p\right\} + o(n^{-2}) \\
&= \frac{p}{2n} + \frac{1}{96n^2}\left\{(3 + \alpha)(7 + 3\alpha)(M - 1) - 6(\alpha + 3)(\alpha + 1)p\right\} + o(n^{-2}). \tag{41}
\end{aligned}
$$

Note that this result could be gained in a more straightforward way, since the risk of m.l.e w.r.t.α-divergence for the multinomial distribution could be expressed in a simple form (see [17] for the straightforward derivation of (41)).

The n^{-1} term is determined by the dimension of the multinomial distribution, while the n^{-2} order term depends only on M once p is fixed. If $(3 + \alpha)(7 + 3\alpha) > 0$, then

Table 1: Approximated $\overset{-1}{ED}(m_1)$ for $B(10, m_1)$

m_1	0.5	0.4	0.3	0.2	0.1	0.01	0.001
$\overset{-1}{ED}(m_1)$	0.0525	0.0526	0.0531	0.0544	0.0584	0.1333	0.8833

n^{-2} order term is a monotonically increasing function of M. When M is minimized, that is when $m_0 = m_1 = \cdots = m_p$, $\overset{\alpha}{ED}(m)$ is minimized. The asymptotically lowest risk among the possible distributions is attained by the equi-probable distribution. The estimation becomes harder as M increases. The term M could be very large when some m_i is close to zero. This justifies the treatment of merging a category of a possibly very low probability with another category for a better inference. The α that is statistically often used such as $\alpha = \pm 1, 0, 3$ satisfies the condition $(3 + \alpha)(7 + 3\alpha) > 0$. However when $-3 < \alpha < -7/3$, these phenomena are vice versa. The equi-probable distribution is the asymptotically highest risk point.

When $\alpha = -1, 0, -3$, we have the following results;

$$\overset{-1}{ED}(m) = \frac{p}{2n} + \frac{1}{24n^2}(2M - 2) + o(n^{-2}), \tag{42}$$

$$\overset{0}{ED}(m) = \frac{p}{2n} + \frac{1}{24n^2}((21/4)M - (9/2)p - 21/4) + o(n^{-2}). \tag{43}$$

$$\overset{-3}{ED}(m) = \frac{p}{2n} + o(n^{-2}), \tag{44}$$

Rather surprisingly, the n^{-2} term for χ^2-divergence vanishes, hence the asymptotic risk up to n^{-2} order is uniform in m.

Figure 1 and Table 1 show the approximated value of $\overset{-1}{ED}(m_1)$ up to the n^{-2} term for the case $p = 1, n = 10$, i.e. $B(10, m_1)$ as m_1 varies. We observe that the risk of the estimation rapidly increases outside of the interval $(0.1, 0.9)$. It is really a hard task to estimate the probability which is less than $1/10$ based on just 10 observations.

$-Example$ $2-$

Second example is the p-dimensional multivariate normal distribution with zero means, that is,

$$X \sim N_p(0, \Sigma), \quad \Sigma = (\sigma_{ij}).$$

The m.l.e. is the sample variance-covariance matrix. Note that if α equals ± 1, the divergence is explicitly given (so called Stein's loss function), hence we can derive the expansion of $\overset{\pm 1}{ED}(\Sigma)$ in a more straightforward way (e.g., for the case $\alpha = -1$, see [17]) .

For this model, we can use the parameter $\sigma_{ij}(1 \le i \le j \le p)$ or $\sigma^{ij}(1 \le i \le j \le p)$, where

$$\sigma^{ij} = (\Sigma^{-1})_{ij}, \quad 1 \le i, j \le p.$$

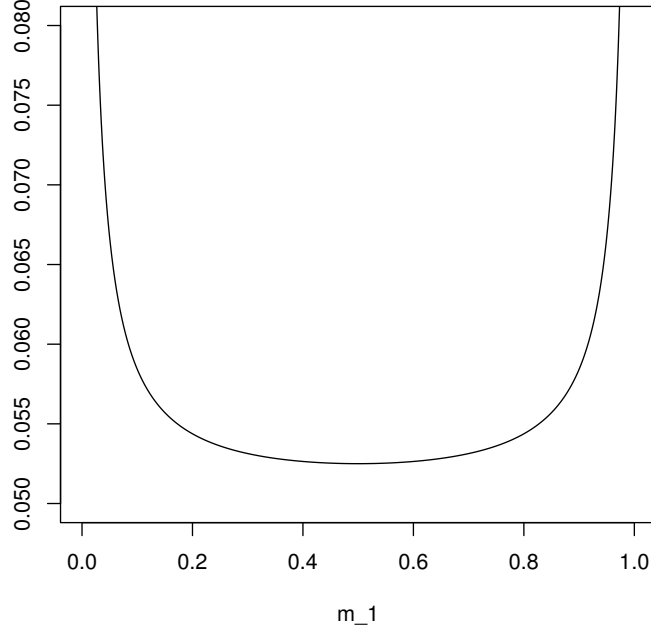We use the notation (i, j), $1 \le i \le j \le p$ to specify the element of the parameters.

Figure 1: Approximated $\overset{-1}{ED}(m_1)$ for $B(10, m_1)$

Since this model is also an exponential family, from (23) and (37), we only have to calculate $T_{ijk}T^{ijk}$, $T_{is}^i T_j^{js}$, $\overset{m}{F}$. These turn out to be as follows (see [17] for the proof) ;

$$
\begin{aligned}
T_{(i,j)(k,l)(s,t)}T^{(i,j)(k,l)(s,t)} &= T_{\sigma_{ij}\sigma_{kl}\sigma_{st}}T^{\sigma_{ij}\sigma_{kl}\sigma_{st}} = T_{\sigma^{ij}\sigma^{kl}\sigma^{st}}T^{\sigma^{ij}\sigma^{kl}\sigma^{st}} \\
&= p^3 + 3p^2 + 4p,
\end{aligned}
\tag{45}
$$

$$
\begin{aligned}
T_{(i,j)(s,t)}^{(i,j)}T_{(k,l)}^{(k,l)(s,t)} &= T_{\sigma_{ij}\sigma_{st}}^{\sigma_{ij}}T_{\sigma_{kl}}^{\sigma_{kl}\sigma_{st}} = T_{\sigma^{ij}\sigma^{st}}^{\sigma^{ij}}T_{\sigma^{kl}}^{\sigma^{kl}\sigma^{st}} \\
&= 2p^3 + 8p^2 + 8p,
\end{aligned}
\tag{46}
$$

$$
\overset{m}{F} = -p^3 - 2p^2 - p.
\tag{47}
$$

Therefore we have

$$
\begin{aligned}
&\overset{\alpha}{ED}(\Sigma) \\
&= \frac{p(p+1)}{4n} \\
&\quad + \frac{1}{24n^2}\left[(\alpha')^2(6p^3 + 30p^2 + 39p) - \alpha'(14p^3 + 48p^2 + 53p) + 10p^3 + 21p^2 + 13p\right] \\
&\quad + o(n^{-2}).
\end{aligned}
\tag{48}
$$

16

Especially when $\alpha = -1, 0, -3$,

$$\overset{-1}{ED}(\Sigma) = \frac{p(p+1)}{4n} + \frac{1}{24n^2}(2p^3 + 3p^2 - p) + o(n^{-2}), \tag{49}$$

$$\overset{0}{ED}(\Sigma) = \frac{p(p+1)}{4n} + \frac{1}{32n^2}(6p^3 + 6p^2 - 5p) + o(n^{-2}). \tag{50}$$

$$\overset{-3}{ED}(\Sigma) = \frac{p(p+1)}{4n} + \frac{1}{8n^2}(2p^3 + 15p^2 + 21p) + o(n^{-2}), \tag{51}$$

Notably $\overset{\alpha}{ED}(\Sigma)$ is not only parameter-invariant but also constant. The risk in estimating the true parameter Σ by m.l.e. is independent of Σ. Actually we have the following lemma.

Lemma 1. *Let $x|\theta$ denote the probability distribution on $\mathfrak{X}$ under the parameter $\theta(\in \Theta)$. Suppose that there exists one to one transformations,*

$$G(x) : \mathfrak{X} \to \mathfrak{X}, \qquad \tilde{G}(\theta) : \Theta \to \Theta$$

satisfying the relation

$$x|\tilde{G}(\theta) \overset{d}{=} G(x)|\theta \text{ or equivalently } G^{-1}(x)|\tilde{G}(\theta) \overset{d}{=} x|\theta. \tag{52}$$

Then $\overset{\alpha}{ED}(\theta) \triangleq E_\theta\big[\overset{\alpha}{D}[\hat{\theta}(\boldsymbol{X}) : \theta]\big]$ is equal to $\overset{\alpha}{ED}(\tilde{G}(\theta)) \triangleq E_{\tilde{G}(\theta)}\big[\overset{\alpha}{D}[\hat{\theta}(\boldsymbol{X}) : \tilde{G}(\theta)]\big]$.

Proof. We use the notation $\overset{\alpha}{D}[x|\theta_1 : x|\theta_2]$ instead of $\overset{\alpha}{D}[\theta_1 : \theta_2]$ for the divergence between the two distributions $x|\theta_1$ and $x|\theta_2$.

$E_\theta\big[\overset{\alpha}{D}[x|\hat{\theta}(\boldsymbol{X}) : x|\theta]\big]$

$= E_\theta\big[\overset{\alpha}{D}[G^{-1}(x)|\tilde{G}(\hat{\theta}(\boldsymbol{X})) : G^{-1}(x)|\tilde{G}(\theta)]\big]$ (because of (52))

$= E_\theta\big[\overset{\alpha}{D}[G^{-1}(x)|\hat{\theta}(G(\boldsymbol{X})) : G^{-1}(x)|\tilde{G}(\theta)]\big]$ (note that $\hat{\theta}(G(\boldsymbol{X})) \overset{d}{=} \tilde{G}(\hat{\theta}(\boldsymbol{X}))$ because of (52))

$= E_{\tilde{G}(\theta)}\big[\overset{\alpha}{D}[G^{-1}(x)|\hat{\theta}(\boldsymbol{X}) : G^{-1}(x)|\tilde{G}(\theta)]\big]$ (because of (52))

$= E_{\tilde{G}(\theta)}\big[\overset{\alpha}{D}[x|\hat{\theta}(\boldsymbol{X}) : x|\tilde{G}(\theta)]\big]$ (because of the invariance property of α-divergence).

$\square$

For arbitrary Σ_1 and Σ_2, if we define

$$G(X) = \Sigma_2^{1/2}\Sigma_1^{-1/2}X, \qquad \tilde{G}(\Sigma) = \Sigma_2^{1/2}\Sigma_1^{-1/2}\Sigma\Sigma_1^{-1/2}\Sigma_2^{1/2},$$

then we have $X|\tilde{G}(\Sigma) \overset{d}{=} G(X)|\Sigma$, hence

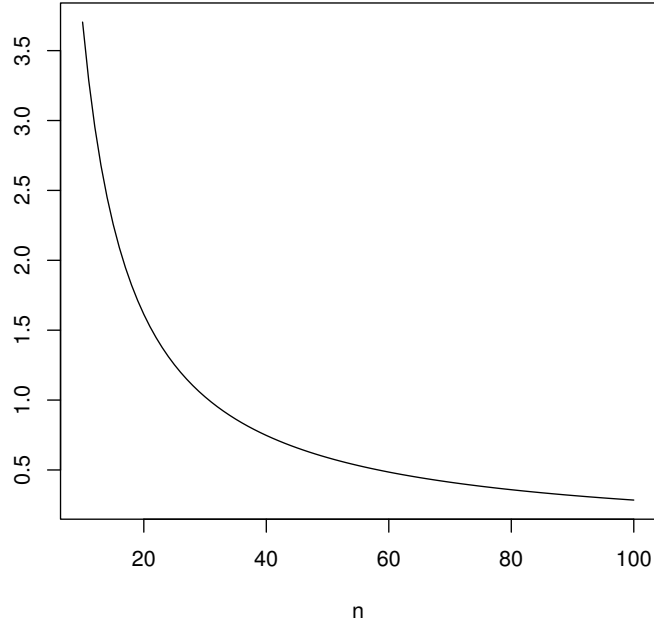$$\overset{\alpha}{ED}(\Sigma_1) = \overset{\alpha}{ED}(\tilde{G}(\Sigma_1)) = \overset{\alpha}{ED}(\Sigma_2).$$

Figure 2: Approximated $\overset{-1}{ED}$ for $N_{10}(0, \Sigma)$

Table 2: Approximated $\overset{-1}{ED}$ for $N_{10}(0, \Sigma)$

n	100	200	300	400	500	800	1000
$\overset{-1}{ED}$	0.2845	0.1399	0.0927	0.0693	0.0554	0.0345	0.0276

We observe "the curse of dimension" in this example. Notice that n^{-1} and n^{-2} terms increase with the second and third power of p respectively. If we increases n and p with the constant ratio p/n, the both terms explode. Figure 2 and Table 2 show the n^{-2}-order-approximated values of $\overset{-1}{ED}$ as n varies, when $p = 10$. When the dimension is 10, we need approximately 500 observations for the same risk as $B(10, 0.5)$ (that is, a 10-times coin toss problem) in estimating the true parameter by m.l.e.

−Example 3−

As a last example, we take a mixture family. For most cases of the mixture family, it is difficult to gain the components of (24) explicitly so that we need to calculate them numerically.

If we use the canonical form (22) with the notation $h_i(x) \triangleq g_i(x) - g_0(x)$, we have

$$l_i = \frac{h_i}{f}, \quad l_{ij} = -\frac{h_i h_j}{f^2} = -l_i l_j,$$

$$l_{ijk} = 2\frac{h_i h_j h_k}{f^3} = 2l_i l_j l_k, \quad l_{ijkl} = -6\frac{h_i h_j h_k h_l}{f^4} = -6l_i l_j l_k l_l.$$

Using these relations, the components of (24) are expressed as follows;

$$\overset{e}{F} = -2g^{ij}g^{ks}L_{ijks} + g^{ks}g^{ui}g^{jk}L_{ijk}L_{sul} + g^{ti}g^{uj}g^{ks}L_{ist}L_{jku},$$

$$T^{ijk}T_{ijk} = L_{ijk}L_{stu}g^{is}g^{jt}g^{ku},$$

$$T_{is}^i T_j^{js} = L_{ijk}L_{stu}g^{ij}g^{st}g^{ku},$$

$$\langle \overset{e}{A}_j^i, \overset{e}{A}_i^j \rangle = g^{jk}g^{li}L_{ikjl} - g^{jk}g^{li}g^{st}L_{iks}L_{jlt} - p,$$

$$\langle \overset{e}{A}_i^i, \overset{e}{A}_j^j \rangle = g^{ik}g^{jl}L_{ikjl} - g^{ik}g^{jl}g^{st}L_{iks}L_{jlt} - p^2.$$

As a more specific example, we consider the mixture of two normal distributions. Let

$$X \sim (1 - \theta_1) * N(0, \sigma^2) + \theta_1 * N(1, \sigma^2), \tag{53}$$

where σ^2 is a known parameter.

We numerically calculated $\overset{e}{F}$, $T^{ijk}T_{ijk}$, $T_{is}^i T_j^{js}$ and $\langle \overset{e}{A}_j^i, \overset{e}{A}_i^j \rangle$ from the above expression by Monte Carlo simulation. We actually calculated those components by generating 10^5 random variables following the mixture distribution (53) under the values of θ_1 from 0.1 to 0.99 by 0.01 increment, from which the n^{-2}-order-approximation of $\overset{-1}{ED}$ was gained. From (24), we notice that

$$\overset{-1}{ED} = \frac{p}{2n} + \frac{1}{24n^2}(-3\overset{e}{F} - T^{ijk}T_{ijk} + 3T_{is}^i T_j^{js} + 12\langle \overset{e}{A}_j^i, \overset{e}{A}_i^j \rangle) + o(n^{-2}).$$

In Figure 3, we can see four U-curves each of which corresponds to the approximated $\overset{-1}{ED}$ of the model (53) with $\sigma^2 = 1/2, 1/5, 1/10$ from the top where $n = 10$ is fixed. The
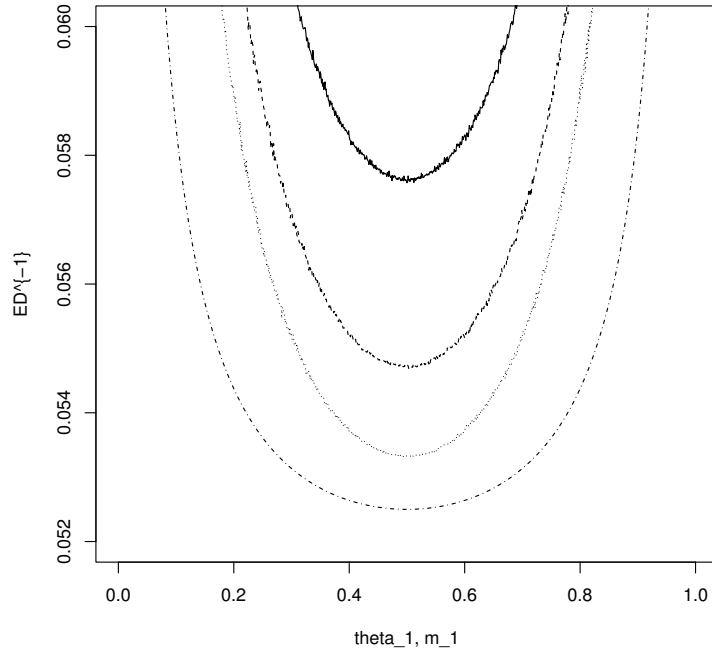
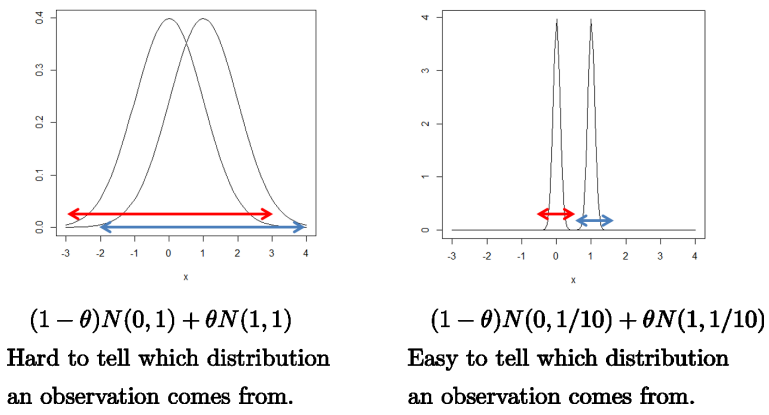Figure 3: Approximated $\widehat{ED}^{-1}$ for the mixture model (53)

$(1-\theta)N(0,1) + \theta N(1,1)$
**Hard to tell which distribution
an observation comes from.**

$(1-\theta)N(0,1/10) + \theta N(1,1/10)$
**Easy to tell which distribution
an observation comes from.**

Figure 4: The relations of two normal distributions

U-curve at the lowest position corresponds to the approximated $\overset{-1}{ED}$ for $B(10,\theta_1)$. It is much harder to specify the model (53) with a large variance compared to the model $B(10,\theta_1)$, since many observations from the model (53) have no trace on whether it comes from $N(0,\sigma^2)$ or $N(1,\sigma^2)$. In contrast, if the variance is small, we can judge from the value of the observation which normal distribution it came from (see Figure 4) . This information helps the inference on θ_1 and the risk in estimating the parameter gets closer to the one for $B(10,\theta_1)$ where head or tail is completely clear.

4 Appendix

4.1 Basic concepts of Information Geometry

Amari [2], Amari and Nagoka [5], Murray and Rice[15] and Calin and Udriște [6] serve as a a general guidance to the information geometry. We only briefly introduce the basic concepts of differential geometry and their concrete forms in the case of statistical manifolds.

We consider an ambient space

$$\mathcal{M} \triangleq \{f(x) \mid f(x) \text{ is a measurable function on } \mathfrak{X}.\}$$

and a scale extension model of $\mathcal{P}$

$$\tilde{\mathcal{P}} \triangleq \{\tilde{f}(x;\tilde{\theta}) \triangleq e^{\theta^0}f(x;\theta) \mid \tilde{\theta} = (\theta^0,\theta) = (\theta^i)_{i=0,1,\dots,p}, \quad \theta \in \Theta, \quad -\infty < \theta^0 < \infty\}.$$

Then $\mathcal{P} \subset \tilde{\mathcal{P}} \subset \mathcal{M}$. We will explain how to construct a Riemmanian manifold structure in $\tilde{\mathcal{P}}$ and $\mathcal{P}$ following the way of Amari ([2]).

We start with $\mathcal{M}$. Consider the variation $g(x;u)$, $-\epsilon < u < \epsilon$ in $\mathcal{M}$ and the corresponding tangent vector ∂_u at $g(x;0)$. The α-representation ($-\infty < \alpha < \infty$) of ∂_u at $g(x;0)$ is defined as

$$\{g(x;0)\}^{-(1+\alpha)/2}\dot{g}(x;0),$$

where

$$\dot{g}(x;0) \triangleq \left. \frac{\partial}{\partial u}g(x;u) \right|_{u=0}.$$

Suppose that another variation $h(x;t)$, $-\epsilon < t < \epsilon$ such as $g(x;0) = h(x;0)$ is given. The inner product between ∂_u and ∂_t at $g(x;0)$ is defined as

$$\begin{aligned}
\langle \partial_u, \partial_t \rangle &\triangleq \int_{\mathfrak{X}} (\alpha\text{-representation of } \partial_u) \times (\alpha\text{-representation of } \partial_t) \times g^\alpha(x;0)d\mu \\
&= \int_{\mathfrak{X}} \{g(x;0)\}^{-1}\dot{g}(x;0)\dot{h}(x;0)d\mu \\
&= \int_{\mathfrak{X}} (\alpha\text{-representation of } \partial_u) \times (-\alpha\text{-representation of } \partial_t)d\mu \\
&= \int_{\mathfrak{X}} g(x;0)\dot{l}g(x;0)\dot{l}h(x;0)d\mu,
\end{aligned}$$

where
$$\dot{l}g(x;0) \triangleq \left.\frac{\partial}{\partial u}\log g(x;u)\right|_{u=0}, \qquad \dot{l}h(x;0) \triangleq \left.\frac{\partial}{\partial t}\log h(x;t)\right|_{t=0}.$$

Let $A(\tilde{\theta})$ be a vector field in $\mathcal{M}$ along $\tilde{\mathcal{P}}$. Its α-representation at $\tilde{\theta}$ is denoted by $\mathring{a}(x;\tilde{\theta})$. Consider a vector $\partial_i \triangleq \partial/\partial\theta^i (i = 0,\dots,p)$. The α-covariant-derivative (α-connection) $(-\infty < \alpha < \infty)$ of A in the space of $\mathcal{M}$ in the direction of ∂_i, $\overset{\alpha}{\nabla}_{\partial_i}A$, is defined as the vector field along $\tilde{\mathcal{P}}$ so that its α-representation at $\tilde{\theta}$ is given by $\partial_i\mathring{a}(x;\tilde{\theta})$.

Now we introduce the geometrical properties of $\tilde{\mathcal{P}}$. First a base of tangent vectors are given by $\partial_i(i = 0,\dots,p)$. The variation of $\tilde{f}(x;\tilde{\theta})$ in $\tilde{\mathcal{P}}$ when θ^i changes gives rise to the tangent vector ∂_i at each $\tilde{\theta}$, and its α-representation is given by

$$\{\tilde{f}(x;\tilde{\theta})\}^{-(1+\alpha)/2}\tilde{f}_i(x;\tilde{\theta}),$$

where $\tilde{f}_i(x;\tilde{\theta}) = \partial_i\tilde{f}(x;\tilde{\theta})$. Components of a Riemannian metric on $\tilde{\mathcal{P}}$ are defined by

$$\begin{aligned}
\tilde{g}_{ij}(\tilde{\theta}) &\triangleq \langle\partial_i,\partial_j\rangle_{\tilde{\theta}} \triangleq \int_{\mathfrak{X}}\{\tilde{f}(x;\tilde{\theta})\}^{-1}\tilde{f}_i(x;\tilde{\theta})\tilde{f}_j(x;\tilde{\theta})d\mu \\
&= \int_{\mathfrak{X}}\tilde{f}(x;\tilde{\theta})\,\tilde{l}_i(x;\tilde{\theta})\,\tilde{l}_j(x;\tilde{\theta})d\mu \\
&= e^{\theta^0}E_\theta[\tilde{l}_i(x;\tilde{\theta})\,\tilde{l}_j(x;\tilde{\theta})],
\end{aligned}$$

where $\tilde{l}_i(x;\tilde{\theta}) \triangleq \partial_i\log\tilde{f}(x;\tilde{\theta})$. Actually $\tilde{g}_{ij}(\tilde{\theta})$ is given by

$$\tilde{g}_{ij}(\tilde{\theta}) = \begin{cases} e^{\theta^0}g_{ij}(\theta) & \text{if } 1 \le i,j \le p, \\ 0 & \text{if } (i,j) = (0,1),\dots,(0,p),(1,0),\dots,(p,0) \\ e^{\theta^0} & \text{if } i = j = 0, \end{cases} \tag{54}$$

where $g_{ij}(\theta) \triangleq \langle\partial_i,\partial_j\rangle_\theta(1 \le i,j \le p)$ is the components of the metric (Fisher information metric) on $\mathcal{P}$ defined by

$$g_{ij}(\theta) \triangleq E_\theta[l_i(x;\theta)\,l_j(x;\theta)]. \tag{55}$$

The second case of (54) indicates $\partial_i \perp \partial_0$ $(i = 1,\dots,p)$, which is derived from

$$\begin{aligned}
&\int_{\mathfrak{X}}\{\tilde{f}(x;\tilde{\theta})\}^{-1}\,\tilde{f}_i(x;\tilde{\theta})\,\tilde{f}_0(x;\tilde{\theta})d\mu \\
&= \int_{\mathfrak{X}}\tilde{f}_i(x;\tilde{\theta})\,d\mu = e^{\theta^0}\int_{\mathfrak{X}}f_i(x;\theta)d\mu = e^{\theta^0}\partial_i\int_{\mathfrak{X}}f(x;\theta)d\mu = 0. \tag{56}
\end{aligned}$$

Another expression of the metric on $\mathcal{P}$

$$g_{ij}(\theta) = -E_\theta[l_{ij}(x;\theta)] \tag{57}$$

is obtained from the relationship

$$E_\theta[l_i(x;\theta)\,l_j(x;\theta) + l_{ij}(x;\theta)] = \int_{\mathfrak{X}}f_{ij}(x;\theta)d\mu = \partial_i\partial_j\int_{\mathfrak{X}}f(x;\theta)d\mu = 0.$$

We use the notation $\tilde{g}^{ij}(\tilde{\theta})$, $g^{ij}(\theta)$ respectively for the components of the inverse matrix of $(\tilde{g}_{ij}(\tilde{\theta}))$, $(g_{ij}(\theta))$.

∂_i $(0 \leq i \leq p)$ is the vector field along $\tilde{\mathcal{P}}$, hence its α-covariant-derivative (α-connection) in the space of $\mathcal{M}$ in the direction of ∂_j $(0 \leq j \leq p)$, $\overline{\overset{\alpha}{\nabla}}_{\partial_j}\partial_i$, could be considered. Its α-representation at $\tilde{\theta}$ is given by

$$\partial_j[\{\tilde{f}(x;\tilde{\theta})\}^{-(1+\alpha)/2}\tilde{f}_i(x;\tilde{\theta})]$$
$$= -\frac{1+\alpha}{2}\{\tilde{f}(x;\tilde{\theta})\}^{-(3+\alpha)/2}\tilde{f}_i(x;\tilde{\theta})\tilde{f}_j(x;\tilde{\theta}) + \{\tilde{f}(x;\tilde{\theta})\}^{-(1+\alpha)/2}\tilde{f}_{ij}(x;\tilde{\theta}).$$

We are mainly concerned with the case $\alpha = \pm 1$. In those cases, more familiar names exist. e-representation and e-covariant-derivative (e-connection) for the case $\alpha = 1$; m-representation and m-covariant-derivative (m-connection) for the case $\alpha = -1$. It turns out that

$$e\text{-representation of } \overline{\overset{e}{\nabla}}_{\partial_j}\partial_i \text{ at } \tilde{\theta}$$
$$= \tilde{l}_{ij}(x;\theta) = \begin{cases} l_{ij}(x;\theta), & \text{if } 1 \leq i,j \leq p, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$m\text{-representation of } \overline{\overset{m}{\nabla}}_{\partial_j}\partial_i \text{ at } \tilde{\theta}$$
$$= \tilde{f}_{ij}(x;\theta) = \begin{cases} e^{\theta^0}f_{ij}(x;\theta), & \text{if } 1 \leq i,j \leq p, \\ e^{\theta^0}f_i(x;\theta), & \text{if } (i,j) = (1,0),\ldots,(p,0), \\ e^{\theta^0}f_j(x;\theta), & \text{if } (i,j) = (0,1),\ldots,(0,p), \\ e^{\theta^0}f(x;\theta), & \text{if } (i,j) = (0,0). \end{cases}$$

Consider the two variations in $\mathcal{M}$,

$$\tilde{f}(x;\tilde{\theta})\exp(ul_{ij}(x;\theta)), \quad |u| < \epsilon, \tag{58}$$
$$\tilde{f}(x;\tilde{\theta}) + te^{\theta^0}f_{ij}(x;\theta), \quad |t| < \epsilon, \tag{59}$$

for $1 \leq i,j \leq p$. ∂_u and ∂_t respectively equals $\overline{\overset{e}{\nabla}}_{\partial_i}\partial_j$ and $\overline{\overset{m}{\nabla}}_{\partial_i}\partial_j$ since the representations coincide.

Let $T_{\tilde{\theta}}\tilde{\mathcal{P}}$ denote the tangent space of $\tilde{\mathcal{P}}$ at $\tilde{\theta}$ (i.e. $\tilde{f}(x;\tilde{\theta})$). Suppose that A is a tangent vector of $\mathcal{M}$ at $\tilde{f}(x;\tilde{\theta})$. The orthogonal projection $\tilde{\pi}$ of A onto $T_{\tilde{\theta}}\tilde{\mathcal{P}}$ is given by

$$\tilde{\pi}(A) = \sum_{0 \leq i,j \leq p} \langle A, \partial_i \rangle \, \tilde{g}^{ij} \, \partial_j.$$

α-covariant-derivative (α-connection) in the space of $\tilde{\mathcal{P}}$ (denoted by $\overset{\approx}{\overset{\alpha}{\nabla}}_{\partial_i}\partial_j, 0 \leq i,j \leq p$) is defined as the orthogonal projection of $\overline{\overset{\alpha}{\nabla}}_{\partial_i}\partial_j$. Actually the following equations hold;

$$\overset{\approx}{\overset{\alpha}{\nabla}}_{\partial_i}\partial_j \triangleq \tilde{\pi}\left(\overline{\overset{\alpha}{\nabla}}_{\partial_i}\partial_j\right)$$

$$= \sum_{0 \le s,t \le p} \langle \overset{\alpha}{\bar{\nabla}}_{\partial_i} \partial_j, \partial_s \rangle \, \tilde{g}^{st} \, \partial_t$$

$$= \sum_{0 \le s,t \le p} \overset{\alpha}{\tilde{\Gamma}}_{ij,s} \, \tilde{g}^{st} \, \partial_t$$

$$= e^{-\theta^0} \sum_{1 \le s,t \le p} \overset{\alpha}{\tilde{\Gamma}}_{ij,s} \, g^{st} \, \partial_t + e^{-\theta^0} \overset{\alpha}{\tilde{\Gamma}}_{ij,0} \, \partial_0$$

where for $0 \le i,j,k \le p$, $-\infty < \alpha < \infty$,

$$\overset{\alpha}{\tilde{\Gamma}}_{ij,k}(\tilde{\theta}) \triangleq \langle \overset{\alpha}{\bar{\nabla}}_{\partial_i} \partial_j, \partial_k \rangle_{\tilde{\theta}}$$

$$= \int_{\mathfrak{X}} (\alpha\text{-representation of } \overset{\alpha}{\bar{\nabla}}_{\partial_i} \partial_j \text{ at } \tilde{\theta}) \times (\alpha\text{-representation of } \partial_k \text{ at } \tilde{\theta})$$

$$\times \tilde{f}^{\alpha}(x; \tilde{\theta}) d\mu$$

$$= -\frac{1+\alpha}{2} \int_{\mathfrak{X}} \tilde{f}_i(x;\tilde{\theta}) \tilde{f}_j(x;\tilde{\theta}) \tilde{f}_k(x;\tilde{\theta}) \tilde{f}^{-2}(x;\tilde{\theta}) d\mu$$

$$+ \int_{\mathfrak{X}} \tilde{f}_{ij}(x;\tilde{\theta}) \tilde{f}_k(x;\tilde{\theta}) \tilde{f}^{-1}(x;\tilde{\theta}) d\mu$$

$$= \frac{1-\alpha}{2} \int_{\mathfrak{X}} \tilde{l}_i(x;\tilde{\theta}) \tilde{l}_j(x;\tilde{\theta}) \tilde{l}_k(x;\tilde{\theta}) \tilde{f}(x;\tilde{\theta}) d\mu$$

$$+ \int_{\mathfrak{X}} \tilde{l}_{ij}(x;\tilde{\theta}) \tilde{l}_k(x;\tilde{\theta}) \tilde{f}(x;\tilde{\theta}) d\mu, \tag{60}$$

where $\tilde{l}_{ij}(x;\theta) \triangleq \partial_i \partial_j \log \tilde{f}(x;\tilde{\theta})$.

The notation $\overset{\alpha}{\tilde{\Gamma}}_{ij,k}$ is called Christoffel's second symbol. We also use Christoffel's first symbol $\overset{\alpha}{\tilde{\Gamma}}{}^k_{ij}$ $(0 \le i,j,k \le p)$ defined by

$$\overset{\alpha}{\tilde{\Gamma}}{}^k_{ij}(\tilde{\theta}) \triangleq \sum_{t=0}^{p} \overset{\alpha}{\tilde{\Gamma}}_{ij,t}(\tilde{\theta}) \, \tilde{g}^{kt}(\tilde{\theta}),$$

$$= \begin{cases} e^{-\theta^0} \sum_{t=1}^{p} \overset{\alpha}{\tilde{\Gamma}}_{ij,t}(\tilde{\theta}) g^{kt}(\theta), & \text{if } 1 \le k \le p, \\ e^{-\theta^0} \overset{\alpha}{\tilde{\Gamma}}_{ij,0}(\tilde{\theta}), & \text{if } k = 0. \end{cases}$$

Cristoffel's symbols are not tensors. For example, when the coordinates are changed from $(i,j,k,\ldots)$ to $(\alpha,\beta,\gamma,\ldots)$, the following exchange rule holds.

$$\overset{\alpha}{\tilde{\Gamma}}_{ij,k} = \overset{\alpha}{\tilde{\Gamma}}_{\alpha\beta,\gamma} B_i^\alpha B_j^\beta B_k^\gamma + B_k^\alpha B_{ij}^\gamma g_{\alpha\gamma} \tag{61}$$

When $\alpha = \pm 1$, Christoffel's symbols are denoted by $\overset{e}{\tilde{\Gamma}}_{ij,k}$, $\overset{e}{\tilde{\Gamma}}{}^k_{ij}$ $(\alpha = 1)$ and $\overset{m}{\tilde{\Gamma}}_{ij,k}$, $\overset{m}{\tilde{\Gamma}}{}^k_{ij}$ $(\alpha = -1)$. The concrete forms of $\overset{e}{\tilde{\Gamma}}_{ij,k}$ and $\overset{m}{\tilde{\Gamma}}_{ij,k}$ are given by

$$\overset{e}{\tilde{\Gamma}}_{ij,k}(\tilde{\theta}) = \int_{\mathfrak{X}} \tilde{l}_{ij}(x;\tilde{\theta}) \tilde{l}_k(x;\tilde{\theta}) \tilde{f}(x;\tilde{\theta}) d\mu, \tag{62}$$

$$\overset{m}{\Gamma}_{ij,k}(\tilde\theta) = \int_{\mathfrak{X}} \left(\tilde l_{ij}(x;\tilde\theta)\tilde l_k(x;\tilde\theta) + \tilde l_i(x;\tilde\theta)\tilde l_j(x;\tilde\theta)\tilde l_k(x;\tilde\theta) \right) \tilde f(x;\tilde\theta)d\mu. \tag{63}$$

From (62) and (63), we notice that

$$\overset{\alpha}{\Gamma}_{ij,k}(\tilde\theta) = \overset{e}{\Gamma}_{ij,k}(\tilde\theta) + \frac{1-\alpha}{2}\left(\overset{m}{\Gamma}_{ij,k}(\tilde\theta) - \overset{e}{\Gamma}_{ij,k}(\tilde\theta) \right) \tag{64}$$

The following relation is an important property for information geometry. It shows the duality between α-covariant derivative and $-\alpha$-covariant derivative. Let A, B, C be vector fields on $\tilde{\mathcal{P}}$, then

$$A\big(\langle B,\ C\rangle\big) = \langle \overset{\alpha}{\tilde\nabla}_A B,\ C\rangle + \langle \overset{-\alpha}{\tilde\nabla}_A C,\ B\rangle. \tag{65}$$

Especially for $0 \le i, j, k \le p$, it turns out that

$$\begin{aligned}
\partial_k \tilde g_{ij} &= \langle \overset{\alpha}{\tilde\nabla}_{\partial_k}\partial_i,\ \partial_j\rangle + \langle \overset{-\alpha}{\tilde\nabla}_{\partial_k}\partial_j,\ \partial_i\rangle \\
&= \langle \overset{\alpha}{\tilde\nabla}_{\partial_k}\partial_i,\ \partial_j\rangle + \langle \overset{-\alpha}{\tilde\nabla}_{\partial_k}\partial_j,\ \partial_i\rangle \\
&= \overset{\alpha}{\Gamma}_{ik,j} + \overset{-\alpha}{\Gamma}_{jk,i}.
\end{aligned} \tag{66}$$

We define α-covariant derivative in the space of $\mathcal{P}$ (denoted by $\overset{\alpha}{\nabla}_{\partial_i}\partial_j, 1 \le i, j \le p$) as the tangent vector field whose value at θ is the orthogonal projection of $\overset{\alpha}{\tilde\nabla}_{\partial_i}\partial_j$ at $(0,\theta)$ on the tangent space of $\mathcal{P}$ at θ. Since $\partial_i \perp \partial_0\ (1 \le i \le p)$, we easily notice that

$$\overset{\alpha}{\nabla}_{\partial_i}\partial_j = \sum_{1\le s,t\le p} \overset{\alpha}{\Gamma}_{ij,s}(\theta)g^{st}(\theta)\partial_t = \sum_{t=1}^{p} \overset{\alpha}{\Gamma}{}_{ij}^{t}(\theta)\partial_t,$$

where $\overset{\alpha}{\Gamma}_{ij,s}(\theta) \triangleq \overset{\alpha}{\Gamma}_{ij,s}((0,\theta))$ and $\overset{\alpha}{\Gamma}{}_{ij}^{t}(\theta) \triangleq \overset{\alpha}{\Gamma}{}_{ij}^{t}((0,\theta))$.

Now we define the second fundamental forms $\overset{\alpha}{A}_{ij}(\tilde\theta)(0 \le i, j \le p, -\infty < \alpha < \infty)$ of $\tilde{\mathcal{P}}$ as

$$\overset{\alpha}{A}_{ij}(\tilde\theta) \triangleq \overset{\alpha}{\tilde\nabla}_{\partial_i}\partial_j - \overset{\alpha}{\nabla}_{\partial_i}\partial_j. \tag{67}$$

Its α-representation, $\overset{\alpha}{a}_{ij}(x;\tilde\theta)$, is given by

$$\begin{aligned}
&\overset{\alpha}{a}_{ij}(x;\tilde\theta) \\
&= (\alpha\text{-representation of } \overset{\alpha}{\tilde\nabla}_{\partial_i}\partial_j) - (\alpha\text{-representation of } \overset{\alpha}{\nabla}_{\partial_i}\partial_j) \\
&= -\frac{1+\alpha}{2}\{\tilde f(x;\tilde\theta)\}^{-(3+\alpha)/2}\tilde f_i(x;\tilde\theta)\tilde f_j(x;\tilde\theta) + \{\tilde f(x;\tilde\theta)\}^{-(1+\alpha)/2}\tilde f_{ij}(x;\tilde\theta) \\
&\quad - e^{-\theta^0}\sum_{1\le s,t\le p} \overset{\alpha}{\Gamma}_{ij,s}(\tilde\theta)\, g^{st}(\theta)\{\tilde f(x;\tilde\theta)\}^{-(1+\alpha)/2}\tilde f_t(x;\tilde\theta) \\
&\quad - e^{-\theta^0}\overset{\alpha}{\Gamma}_{ij,0}(\tilde\theta)\{\tilde f(x;\tilde\theta)\}^{-(1+\alpha)/2}\tilde f_0(x;\tilde\theta).
\end{aligned}$$

For the case $\alpha = \pm 1$, we use the notations $\overset{e}{A}_{ij}(\tilde{\theta})$, $\overset{e}{a}_{ij}(x;\tilde{\theta})$ ($\alpha = 1$) and $\overset{m}{A}_{ij}(\tilde{\theta})$, $\overset{m}{a}_{ij}(x;\tilde{\theta})$ ($\alpha = -1$). The concrete forms of $\overset{e}{a}_{ij}(x;\tilde{\theta})$ and $\overset{m}{a}_{ij}(x;\tilde{\theta})$ are given as follows;

$$\overset{e}{a}_{ij}(x;\tilde{\theta}) = \tilde{l}_{ij}(x;\tilde{\theta}) - e^{-\theta^0} \sum_{1 \le t \le p} \overset{e}{\Gamma}{}^t_{ij}(\tilde{\theta})\, \tilde{l}_t(x;\tilde{\theta}) + e^{-\theta^0} \tilde{g}_{ij}(\tilde{\theta}),$$

$$\overset{m}{a}_{ij}(x;\tilde{\theta}) = \tilde{f}_{ij}(x;\tilde{\theta}) - e^{-\theta^0} \sum_{1 \le t \le p} \overset{m}{\Gamma}{}^t_{ij}(\tilde{\theta})\, \tilde{f}_t(x;\tilde{\theta}) - e^{-\theta^0} \tilde{f}(x;\tilde{\theta}) \int_{\mathfrak{X}} \tilde{f}_{ij}(x;\tilde{\theta})d\mu.$$

Especially when $1 \le i,j \le p$

$$\overset{e}{a}_{ij}(x;\tilde{\theta}) = l_{ij}(x;\theta) - e^{-\theta^0} \sum_{1 \le t \le p} \overset{e}{\Gamma}{}^t_{ij}(\tilde{\theta})\, l_t(x;\theta) + g_{ij}(\theta), \tag{68}$$

$$\overset{m}{a}_{ij}(x;\tilde{\theta}) = \tilde{f}_{ij}(x;\tilde{\theta}) - e^{-\theta^0} \sum_{1 \le t \le p} \overset{m}{\Gamma}{}^t_{ij}(\tilde{\theta})\, \tilde{f}_t(x;\tilde{\theta}), \tag{69}$$

since

$$\int_{\mathfrak{X}} \tilde{f}_{ij}(x;\tilde{\theta})d\mu = e^{\theta^0} \int_{\mathfrak{X}} f_{ij}(x;\theta)d\mu = e^{\theta^0} \partial_i \partial_j \int_{\mathfrak{X}} f(x;\theta)d\mu = 0.$$

From the definition of the covariant derivative, the α-representation of $\overset{\alpha}{\overline{\nabla}}_{\partial_h} \overset{\alpha}{A}_{ij}$ is given by $\partial_h \overset{\alpha}{a}_{ij}(x;\tilde{\theta})$. The following equation holds;

$$\int_{\mathfrak{X}} \left(\partial_h \overset{\alpha}{a}_{ij}(x;\tilde{\theta}) \right) \tilde{f}_k(x;\tilde{\theta})\{\tilde{f}(x;\tilde{\theta})\}^{(\alpha-1)/2}d\mu = \left\langle \overset{\alpha}{\overline{\nabla}}_{\partial_h} \overset{\alpha}{A}_{ij}, \partial_k \right\rangle_{\tilde{\theta}}$$

$$= -\left\langle \overset{\alpha}{A}_{ij}, \overset{-\alpha}{\overline{\nabla}}_{\partial_h} \partial_k \right\rangle_{\tilde{\theta}}$$

$$= -\left\langle \overset{\alpha}{A}_{ij}, \overset{-\alpha}{\overline{\nabla}}_{\partial_h} \partial_k - \tilde{\pi}\left(\overset{-\alpha}{\overline{\nabla}}_{\partial_h} \partial_k \right) \right\rangle_{\tilde{\theta}}$$

$$= -\left\langle \overset{\alpha}{A}_{ij}, \overset{-\alpha}{A}_{hk} \right\rangle_{\tilde{\theta}}, \tag{70}$$

where the second and third equations hold since $\langle \overset{\alpha}{A}_{ij}, \partial_k \rangle_{\tilde{\theta}} = 0$.

We often use another type of second fundamental form

$$\overset{\alpha}{A}{}^j_i(\tilde{\theta}) \triangleq \overset{\alpha}{A}_{ik}(\tilde{\theta})g^{kj}(\tilde{\theta}) = \overset{\alpha}{A}_{ki}(\tilde{\theta})g^{kj}(\tilde{\theta}). \tag{71}$$

The simplified notations $\overset{\alpha}{A}_{ij}(\theta)$, $\overset{\alpha}{A}{}^j_i(\theta)$ are also used instead of $\overset{\alpha}{A}_{ij}((0,\theta))$ and $\overset{\alpha}{A}{}^j_i((0,\theta))$.

Lastly, we refer to Riemannian curvature tensor of $\mathcal{P}$. For vector fields A, B, C on $\mathcal{P}$, Riemannian curvature tensor with respect to $\overset{\alpha}{\nabla}$ is defined as

$$\overset{\alpha}{R}(A,B)\,C \triangleq \overset{\alpha}{\nabla}_A \left(\overset{\alpha}{\nabla}_B C \right) - \overset{\alpha}{\nabla}_B \left(\overset{\alpha}{\nabla}_A C \right) - \overset{\alpha}{\nabla}_{AB-BA} C.$$

The components of Riemannian curvature tensor, $\overset{\alpha}{R}_{ijk}{}^l(\theta)$, are defined as the unique function on $\mathcal{P}$ which satisfies

$$\overset{\alpha}{R}(\partial_i, \partial_j)\,\partial_k = \sum_{l=1}^p \overset{\alpha}{R}_{ijk}{}^l(\theta)\partial_l.$$

More specifically it is given by

$$\overset{\alpha}{R}_{ijk}{}^{l}(\theta) = \partial_i \overset{\alpha}{\Gamma}{}_{jk}^{l}(\theta) - \partial_j \overset{\alpha}{\Gamma}{}_{ik}^{l}(\theta) + \overset{\alpha}{\Gamma}{}_{ir}^{l}(\theta)\overset{\alpha}{\Gamma}{}_{jk}^{r}(\theta) - \overset{\alpha}{\Gamma}{}_{jr}^{l}(\theta)\overset{\alpha}{\Gamma}{}_{ik}^{r}(\theta). \tag{72}$$

4.2 Geometric Interpretation of Derivatives of Log-likelihood

The expectation of the derivatives of the log-likelihood $l(x;\theta)$ can be expressed in the geometric terms introduced in the previous subsection.

Lemma 2. *For $1 \le i,\ j,\ h,\ k \le p$, The following relations hold.*

$$E_\theta[l_i(x;\theta)] = 0, \tag{73}$$

$$E_\theta[l_i(x;\theta)l_j(x;\theta)] = -E_\theta[l_{ij}(x;\theta)] = g_{ij}(\theta), \tag{74}$$

$$E_\theta[l_{ij}(x;\theta)l_k(x;\theta)] = \overset{e}{\Gamma}_{ij,k}(\theta), \tag{75}$$

$$E_\theta[l_i(x;\theta)l_j(x;\theta)l_k(x;\theta)] = \overset{m}{\Gamma}_{ij,k}(\theta) - \overset{e}{\Gamma}_{ij,k}(\theta), \tag{76}$$

$$E_\theta[l_{ijk}(x;\theta)] = -\big(\overset{e}{\Gamma}_{ij,k}(\theta) + \overset{e}{\Gamma}_{ik,j}(\theta) + \overset{m}{\Gamma}_{jk,i}(\theta)\big), \tag{77}$$

$$E_\theta[l_{ijh}(x;\theta)l_k(x;\theta)] = \sum_{t=1}^{p}\big(\partial_h \overset{e}{\Gamma}{}_{ij}^{t}(\theta)\big)g_{tk}(\theta) + \sum_{t=1}^{p}\overset{e}{\Gamma}{}_{ij}^{t}(\theta)\overset{e}{\Gamma}_{th,k}(\theta)$$
$$- \langle \overset{e}{A}_{ij}(\theta), \overset{m}{A}_{hk}(\theta)\rangle, \tag{78}$$

$$E_\theta[l_{ij}(x;\theta)l_{kh}(x;\theta)] = \langle \overset{e}{A}_{ij}(\theta), \overset{e}{A}_{kh}(\theta)\rangle + \sum_{t=1}^{p}\overset{e}{\Gamma}{}_{ij}^{t}(\theta)\overset{e}{\Gamma}_{kh,t}(\theta)$$
$$+ g_{ij}(\theta)g_{kh}(\theta), \tag{79}$$

$$E_\theta[l_{ij}(x;\theta)l_k(x;\theta)l_h(x;\theta)] = \langle \overset{e}{A}_{ij}(\theta), \big(\overset{m}{A}_{kh}(\theta) - \overset{e}{A}_{kh}(\theta)\big)\rangle - g_{ij}(\theta)g_{kh}(\theta)$$
$$+ \sum_{s=1}^{p}\overset{e}{\Gamma}_{ij,s}(\theta)\big(\overset{m}{\Gamma}{}_{kh}^{s}(\theta) - \overset{e}{\Gamma}{}_{kh}^{s}(\theta)\big), \tag{80}$$

$$E_\theta[l_{ijkh}(x;\theta)] = -\big(\partial_k \overset{e}{\Gamma}_{ij,h}(\theta) + \partial_k \overset{e}{\Gamma}_{ih,j}(\theta) + \partial_k \overset{m}{\Gamma}_{jh,i}(\theta)\big)$$
$$- \sum_{t=1}^{p}\big(\partial_h \overset{e}{\Gamma}{}_{ij}^{t}(\theta)\big)g_{tk}(\theta) - \sum_{t=1}^{p}\overset{e}{\Gamma}{}_{ij}^{t}(\theta)\overset{e}{\Gamma}_{th,k}(\theta)$$
$$+ \langle \overset{e}{A}_{ij}(\theta), \overset{m}{A}_{hk}(\theta)\rangle. \tag{81}$$

- *Proof* -
Equation (74) is the result already mentioned in (55) and (57). (75) and (76) are special cases of (62) and (63) when $\tilde\theta = (0,\theta)$ and $1 \le i,j,k, \le p$. In the following proof, we will abbreviate $f(x;\theta), l_i(x;\theta), l_{ij}(x;\theta), \cdots$ respectively to $f, l_i, l_{ij}, \cdots$.
- *Proof of* (73)-
$$E_\theta[l_i] = \int_{\mathfrak{x}} f_i(x;\theta)d\mu = \partial_i \int_{\mathfrak{x}} f(x;\theta)d\mu = 0.$$

- *Proof of (77)*-
By differentiating the both sides of $l_{ij} = -l_i l_j + f_{ij}/f$, we get the following equation.

$$l_{ijk} = -l_{ik} l_j - l_i l_{jk} + \frac{f_{ijk}}{f} - \frac{f_{ij} f_k}{f^2}$$
$$= -l_{ik} l_j - l_i l_{jk} + \frac{f_{ijk}}{f}$$
$$- \left(\frac{f_{ij}}{f} - \frac{f_i}{f} \frac{f_j}{f} + \frac{f_i}{f} \frac{f_j}{f} \right) \left(\frac{f_k}{f} \right)$$
$$= -(l_{ij} l_k + l_{ik} l_j + l_{jk} l_i) - l_i l_j l_k + \frac{f_{ijk}}{f}.$$

Take expectation of both sides of the above equation, and use (75), (76) and the relation

$$E_\theta[f_{ijk}/f] = \int_{\mathfrak{X}} f_{ijk} d\mu = \partial_i \partial_j \partial_k \int_{\mathfrak{X}} f d\mu = 0,$$

then we obtain the results.
- *Proof of (78)*-
From (68) for the case $\tilde{\theta} = (0, \theta)$,

$$l_{ij}(x; \theta) = \overset{e}{a}_{ij}(x; \theta) + \sum_{t=1}^{p} \overset{e}{\Gamma}{}_{ij}^{t}(\theta) l_t(x; \theta) - g_{ij}(\theta). \tag{82}$$

By differentiating both sides of this equation, we have

$$l_{ijh}(x; \theta) = \partial_h \overset{e}{a}_{ij}(x; \theta) + \sum_{t=1}^{p} (\partial_h \overset{e}{\Gamma}{}_{ij}^{t}(\theta)) l_t(x; \theta) + \sum_{t=1}^{p} \overset{e}{\Gamma}{}_{ij}^{t}(\theta) l_{th}(x; \theta) - \partial_h g_{ij}(\theta).$$

Consequently

$$E_\theta[l_{ijh} l_k]$$
$$= E_\theta[(\partial_h \overset{e}{a}_{ij}) l_k] + \sum_{t=1}^{p} (\partial_h \overset{e}{\Gamma}{}_{ij}^{t}(\theta)) E_\theta[l_t l_k] + \sum_{t=1}^{p} \overset{e}{\Gamma}{}_{ij}^{t}(\theta) E_\theta[l_{th} l_k] - (\partial_h g_{ij}(\theta)) E_\theta[l_k]$$
$$= -\langle \overset{e}{A}_{ij}, \overset{m}{A}_{hk} \rangle_\theta + \sum_{t=1}^{p} (\partial_h \overset{e}{\Gamma}{}_{ij}^{t}(\theta)) g_{tk}(\theta) + \sum_{t=1}^{p} \overset{e}{\Gamma}{}_{ij}^{t}(\theta) \overset{e}{\Gamma}{}_{th,k}(\theta),$$

where the second equation comes from (70) (for the case $\tilde{\theta} = (0, \theta)$ and $\alpha = 1$), (73), (74) and (75).
- *Proof of (79)*-
From the definition of the second fundamental form, it turns out that $\langle \overset{e}{A}_{ij}, \partial_t \rangle = 0$ ($0 \leq i, j, t \leq p$), hence, for $1 \leq i, j, t \leq p$,

$$E_\theta[\overset{e}{a}_{ij} l_t] = E_\theta[\overset{e}{a}_{ij}] = 0.$$

28

Using (82) for l_{ij}, l_{kh} and the relation (73), (74), we notice the following equation holds.

$$E_\theta[l_{ij}l_{kh}] = E_\theta[\overset{e}{a}_{ij}\overset{e}{a}_{kh}] + \sum_{1 \le s,t \le p} \overset{e}{\Gamma}{}^t_{ij} \overset{e}{\Gamma}{}^s_{kh} E_\theta[l_t l_s] + g_{ij}g_{kh}$$

$$= \langle \overset{e}{A}_{ij}, \overset{e}{A}_{kh} \rangle + \sum_{t=1}^p \overset{e}{\Gamma}{}^t_{ij} \overset{e}{\Gamma}_{kh,t} + g_{ij}g_{kh}.$$

- *Proof of* (80)-
Since $l_k l_h = f_{kh}/f - l_{kh}$,

$$E_\theta[l_{ij}l_k l_h] = \int_{\mathfrak{X}} l_{ij}(f_{kh}/f - l_{kh})f d\mu = \int_{\mathfrak{X}} l_{ij}f_{kh}d\mu - E_\theta[l_{ij}l_{kh}]. \tag{83}$$

From (69) for the case $\tilde{\theta} = (0, \theta)$, we have

$$f_{kh}(x;\theta) = \overset{m}{a}_{kh}(x;\theta) + \sum_{t=1}^p \overset{m}{\Gamma}{}^t_{kh}(\theta) \, f_t(x;\theta). \tag{84}$$

If we substitute l_{ij} and f_{kh} in the integrand of (83) with the right-hand sides of (82) and (84) respectively, we have

$$\int_{\mathfrak{X}} l_{ij}f_{kh}d\mu$$

$$= \int_{\mathfrak{X}} \left(\overset{e}{a}_{ij}\overset{m}{a}_{kh} + \overset{e}{a}_{ij} \sum_{t=1}^p \overset{m}{\Gamma}{}^t_{kh} f_t + \overset{m}{a}_{kh} \sum_{t=1}^p \overset{e}{\Gamma}{}^t_{ij} l_t \right.$$

$$\left. + \sum_{t=1}^p \overset{e}{\Gamma}{}^t_{ij} l_t \sum_{s=1}^p \overset{m}{\Gamma}{}^s_{kh} f_s - g_{ij}\overset{m}{a}_{kh} - g_{ij} \sum_{t=1}^p \overset{m}{\Gamma}{}^t_{kh} f_t \right) d\mu$$

$$= \langle \overset{e}{A}_{ij}, \overset{m}{A}_{kh} \rangle + \sum_{t=1}^p \overset{m}{\Gamma}{}^t_{kh} \langle \overset{e}{A}_{ij}, \partial_t \rangle + \sum_{t=1}^p \overset{e}{\Gamma}{}^t_{ij} \langle \partial_t, \overset{m}{A}_{kh} \rangle$$

$$+ \sum_{1 \le s,t \le p} \overset{m}{\Gamma}{}^s_{kh} \overset{e}{\Gamma}{}^t_{ij} g_{st} - g_{ij}\langle \overset{m}{A}_{kh}, \partial_0 \rangle - g_{ij} \sum_{t=1}^p \overset{m}{\Gamma}{}^t_{kh} \langle \partial_t, \partial_0 \rangle$$

$$= \langle \overset{e}{A}_{ij}, \overset{m}{A}_{kh} \rangle + \sum_{s=1}^p \overset{m}{\Gamma}{}^s_{kh} \overset{e}{\Gamma}_{ij,s}.$$

since $\langle \overset{e}{A}_{ij}, \partial_k \rangle = \langle \overset{m}{A}_{ij}, \partial_k \rangle = 0$ for $0 \le i, j, k \le p$, and $\langle \partial_t, \partial_0 \rangle = 0$ for $1 \le t \le p$. Combine this equation with (79) and (83), then we obtain the results.
- *Proof of* (81)-
If we differentiate both sides of the equation

$$\int_{\mathfrak{X}} l_{ijh}(x;\theta)f(x;\theta)d\mu = -(\overset{e}{\Gamma}_{ij,h}(\theta) + \overset{e}{\Gamma}_{ih,j}(\theta) + \overset{m}{\Gamma}_{jh,i}(\theta)) \quad (\text{see } (77)),$$

then we have

$$\partial_k \int_{\mathfrak{X}} l_{ijh}(x;\theta)f(x;\theta)d\mu = -(\partial_k \overset{e}{\Gamma}_{ij,h}(\theta) + \partial_k \overset{e}{\Gamma}_{ih,j}(\theta) + \partial_k \overset{m}{\Gamma}_{jh,i}(\theta))$$

The left hand side equals $E_\theta[l_{ijkh}] + E_\theta[l_{ijh}l_k]$. From (78), we have

$$E_\theta[l_{ijkh}] = -(\partial_k \overset{e}{\Gamma}_{ij,h} + \partial_k \overset{e}{\Gamma}_{ih,j} + \partial_k \overset{m}{\Gamma}_{jh,i}) - E_\theta[l_{ijh}l_k]$$

$$= -(\partial_k \overset{e}{\Gamma}_{ij,h} + \partial_k \overset{e}{\Gamma}_{ih,j} + \partial_k \overset{m}{\Gamma}_{jh,i})$$

$$- \sum_{t=1}^p (\partial_h \overset{e}{\Gamma}_{ij}^t)g_{tk} - \sum_{t=1}^p \overset{e}{\Gamma}_{ij}^t \overset{e}{\Gamma}_{th,k} + \langle \overset{e}{A}_{ij}, \overset{m}{A}_{hk} \rangle.$$

4.3 Expansion of Divergence

Substitute θ_1 and θ_2 in (2) respectively with θ and θ_0. Fix θ_0 in $\overset{\alpha}{D}[\theta : \theta_0]$ and treat it as the function of θ. Then Taylor expansion of $\overset{\alpha}{D}(\theta) = \overset{\alpha}{D}[\theta : \theta_0]$ around θ_0 is given by

$$\overset{\alpha}{D}(\theta)$$

$$= \sum_{i=1}^p \left(\epsilon_i \overset{\alpha}{D}[\theta_0 : \theta_0]\right)(\theta^i - \theta_0^i) + \frac{1}{2}\sum_{i=1}^p \sum_{j=1}^p \left(\epsilon_i \epsilon_j \overset{\alpha}{D}[\theta_0 : \theta_0]\right)(\theta^i - \theta_0^i)(\theta^j - \theta_0^j)$$

$$+ \frac{1}{6}\sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \left(\epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta_0 : \theta_0]\right)(\theta^i - \theta_0^i)(\theta^j - \theta_0^j)(\theta^k - \theta_0^k)$$

$$+ \frac{1}{24}\sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \left(\epsilon_i \epsilon_j \epsilon_k \epsilon_l \overset{\alpha}{D}[\theta_0 : \theta_0]\right)(\theta^i - \theta_0^i)(\theta^j - \theta_0^j)(\theta^k - \theta_0^k)(\theta^l - \theta_0^l)$$

$$+ O(||\theta - \theta_0||^5), \tag{85}$$

where ϵ_i is the partial differentiation of $\overset{\alpha}{D}[\theta_1 : \theta_2]$ with respect to θ_1^i.

According to Eguchi's relationship (see [8]),

$$\epsilon_i \overset{\alpha}{D}[\theta_0 : \theta_0] = 0, \tag{86}$$

$$\epsilon_i \epsilon_j \overset{\alpha}{D}[\theta_0 : \theta_0] = g_{ij}(\theta_0), \tag{87}$$

$$\epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta_0 : \theta_0] = \overset{\alpha}{\Gamma}_{ij,k}(\theta_0) + \overset{\alpha}{\Gamma}_{ik,j}(\theta_0) + \overset{-\alpha}{\Gamma}_{kj,i}(\theta_0), \tag{88}$$

where g_{ij} and $\overset{\alpha}{\Gamma}_{ij,k}$ are respectively the components of Fisher information metric and α-connection of Riemannian manifold $\mathcal{P}$ (see Section 4.1).

In this section we reveal the geometrical meaning of the forth derivative term

$$\epsilon_i \epsilon_j \epsilon_k \epsilon_l \overset{\alpha}{D}[\theta_0 : \theta_0]$$

(see Eguchi [9] as a related work). Let $\delta_i(i = 1, \ldots, p)$ denote the partial derivative w.r.t. θ_2^i of $\overset{\alpha}{D}[\theta_1 : \theta_2]$. From (88) and (64), we have

$$
\begin{aligned}
\epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta : \theta] &= \overset{\alpha}{\Gamma}_{ij,k}(\theta) + \overset{\alpha}{\Gamma}_{ik,j}(\theta) + \overset{-\alpha}{\Gamma}_{kj,i}(\theta) \\
&= \overset{e}{\Gamma}_{ij,k}(\theta) + \overset{e}{\Gamma}_{ik,j}(\theta) + \overset{e}{\Gamma}_{kj,i}(\theta) \\
&\quad + \frac{3-\alpha}{2}\big(\overset{m}{\Gamma}_{ij,k}(\theta) - \overset{e}{\Gamma}_{ij,k}(\theta)\big) \text{ (see (64))}.
\end{aligned} \tag{89}
$$

Differentiate both sides of this equation in θ^l, then we have

$$
\begin{aligned}
\epsilon_i \epsilon_j \epsilon_k \epsilon_l \overset{\alpha}{D}[\theta : \theta] &= -\delta_l \epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta : \theta] + \partial_l \overset{e}{\Gamma}_{ij,k}(\theta) + \partial_l \overset{e}{\Gamma}_{ik,j}(\theta) + \partial_l \overset{e}{\Gamma}_{kj,i}(\theta) \\
&\quad + \frac{3-\alpha}{2}\partial_l \big(\overset{m}{\Gamma}_{ij,k} - \overset{e}{\Gamma}_{ij,k}\big).
\end{aligned} \tag{90}
$$

From (2), we have

$$
\begin{aligned}
&\delta_l \epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta_1 : \theta_2] \\
&= -\frac{(1-\alpha)^2}{4} \int_{\mathfrak{x}} l_i(x;\theta_1) l_j(x;\theta_1) l_k(x;\theta_1) l_l(x;\theta_2) f^{(1-\alpha)/2}(x;\theta_1) f^{(1+\alpha)/2}(x;\theta_2) d\mu \\
&\quad - \frac{1-\alpha}{2} \int_{\mathfrak{x}} l_{ik}(x;\theta_1) l_j(x;\theta_1) l_l(x;\theta_2) f^{(1-\alpha)/2}(x;\theta_1) f^{(1+\alpha)/2}(x;\theta_2) d\mu \\
&\quad - \frac{1-\alpha}{2} \int_{\mathfrak{x}} l_i(x;\theta_1) l_{jk}(x;\theta_1) l_l(x;\theta_2) f^{(1-\alpha)/2}(x;\theta_1) f^{(1+\alpha)/2}(x;\theta_2) d\mu \\
&\quad - \frac{1-\alpha}{2} \int_{\mathfrak{x}} l_{ij}(x;\theta_1) l_k(x;\theta_1) l_l(x;\theta_2) f^{(1-\alpha)/2}(x;\theta_1) f^{(1+\alpha)/2}(x;\theta_2) d\mu \\
&\quad - \int_{\mathfrak{x}} l_{ijk}(x;\theta_1) l_l(x;\theta_2) f^{(1-\alpha)/2}(x;\theta_1) f^{(1+\alpha)/2}(x;\theta_2) d\mu.
\end{aligned}
$$

Consequently we have

$$
\begin{aligned}
\delta_l \epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta : \theta] &= -\frac{(1-\alpha)^2}{4} E_\theta[l_i l_j l_k l_l] \\
&\quad - \frac{1-\alpha}{2}\Big\{ E_\theta[l_{ik} l_j l_l] + E_\theta[l_{jk} l_i l_l] + E_\theta[l_{ij} l_k l_l] \Big\} \\
&\quad - E_\theta[l_{ijk} l_l].
\end{aligned} \tag{91}
$$

The relation

$$
\begin{aligned}
\partial_l E_\theta[l_i l_j l_k] &= \partial_l \int_{\mathfrak{x}} l_i l_j l_k f d\mu \\
&= \int_{\mathfrak{x}} l_{il} l_j l_k f d\mu + \int_{\mathfrak{x}} l_i l_{jl} l_k f d\mu + \int_{\mathfrak{x}} l_i l_j l_{kl} d\mu + \int_{\mathfrak{x}} l_i l_j l_k l_l f d\mu \\
&= E_\theta[l_{il} l_j l_k] + E_\theta[l_i l_{jl} l_k] + E_\theta[l_i l_j l_{kl}] + E_\theta[l_i l_j l_k l_l],
\end{aligned}
$$

31

and (76) lead to

$$E_\theta[l_i l_j l_k l_l] = \partial_l \big(\overset{m}{\Gamma}_{ij,k} - \overset{e}{\Gamma}_{ij,k} \big) - \big(E_\theta[l_{il} l_j l_k] + E_\theta[l_i l_{jl} l_k] + E_\theta[l_i l_j l_{kl}] \big). \tag{92}$$

Substitute (92), (78) and (80) into (91), and let α' denote $(1-\alpha)/2$, Then we have

$$\delta_l \epsilon_i \epsilon_j \epsilon_k \overset{\alpha}{D}[\theta : \theta]$$

$$= -(\alpha')^2 \partial_l \overset{m}{\Gamma}_{ij,k} + (\alpha')^2 \partial_l \overset{e}{\Gamma}_{ij,k}$$

$$+ (\alpha')^2 \Big(\langle \overset{e}{A}_{il}, \overset{m}{A}_{jk} \rangle - \langle \overset{e}{A}_{il}, \overset{e}{A}_{jk} \rangle - g_{il} g_{jk} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{il,s} \overset{m}{\Gamma}_{jk}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{il,s} \overset{e}{\Gamma}_{jk}^{s}$$

$$+ \langle \overset{e}{A}_{jl}, \overset{m}{A}_{ik} \rangle - \langle \overset{e}{A}_{jl}, \overset{e}{A}_{ik} \rangle - g_{jl} g_{ik} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{jl,s} \overset{m}{\Gamma}_{ik}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{jl,s} \overset{e}{\Gamma}_{ik}^{s}$$

$$+ \langle \overset{e}{A}_{kl}, \overset{m}{A}_{ij} \rangle - \langle \overset{e}{A}_{kl}, \overset{e}{A}_{ij} \rangle - g_{kl} g_{ij} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{kl,s} \overset{m}{\Gamma}_{ij}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{kl,s} \overset{e}{\Gamma}_{ij}^{s} \Big)$$

$$- \alpha' \Big(\langle \overset{e}{A}_{ik}, \overset{m}{A}_{jl} \rangle - \langle \overset{e}{A}_{ik}, \overset{e}{A}_{jl} \rangle - g_{ik} g_{jl} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{ik,s} \overset{m}{\Gamma}_{jl}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{ik,s} \overset{e}{\Gamma}_{jl}^{s}$$

$$+ \langle \overset{e}{A}_{jk}, \overset{m}{A}_{il} \rangle - \langle \overset{e}{A}_{jk}, \overset{e}{A}_{il} \rangle - g_{jk} g_{il} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{jk,s} \overset{m}{\Gamma}_{il}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{jk,s} \overset{e}{\Gamma}_{il}^{s}$$

$$+ \langle \overset{e}{A}_{ij}, \overset{m}{A}_{kl} \rangle - \langle \overset{e}{A}_{ij}, \overset{e}{A}_{kl} \rangle - g_{ij} g_{kl} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{ij,s} \overset{m}{\Gamma}_{kl}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{ij,s} \overset{e}{\Gamma}_{kl}^{s} \Big)$$

$$- \sum_{t=1}^{p} \big(\partial_k \overset{e}{\Gamma}_{ij}^{t} \big) g_{tl} - \sum_{t=1}^{p} \overset{e}{\Gamma}_{ij}^{t} \overset{e}{\Gamma}_{tk,l} + \langle \overset{e}{A}_{ij}, \overset{m}{A}_{kl} \rangle,$$

where the definition and the geometrical meaning of each notation are described in Section 4.1. From this equation and (90), we have

$$\epsilon_i \epsilon_j \epsilon_k \epsilon_l \overset{\alpha}{D}[\theta : \theta]$$

$$= (\alpha')^2 \partial_l \overset{m}{\Gamma}_{ij,k} - (\alpha')^2 \partial_l \overset{e}{\Gamma}_{ij,k}$$

$$- (\alpha')^2 \Big(\langle \overset{e}{A}_{il}, \overset{m}{A}_{jk} \rangle - \langle \overset{e}{A}_{il}, \overset{e}{A}_{jk} \rangle - g_{il} g_{jk} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{il,s} \overset{m}{\Gamma}_{jk}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{il,s} \overset{e}{\Gamma}_{jk}^{s}$$

$$+ \langle \overset{e}{A}_{jl}, \overset{m}{A}_{ik} \rangle - \langle \overset{e}{A}_{jl}, \overset{e}{A}_{ik} \rangle - g_{jl} g_{ik} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{jl,s} \overset{m}{\Gamma}_{ik}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{jl,s} \overset{e}{\Gamma}_{ik}^{s}$$

$$+ \langle \overset{e}{A}_{kl}, \overset{m}{A}_{ij} \rangle - \langle \overset{e}{A}_{kl}, \overset{e}{A}_{ij} \rangle - g_{kl} g_{ij} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{kl,s} \overset{m}{\Gamma}_{ij}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{kl,s} \overset{e}{\Gamma}_{ij}^{s} \Big)$$

$$+ \alpha' \Big(\langle \overset{e}{A}_{ik}, \overset{m}{A}_{jl} \rangle - \langle \overset{e}{A}_{ik}, \overset{e}{A}_{jl} \rangle - g_{ik} g_{jl} + \sum_{s=1}^{p} \overset{e}{\Gamma}_{ik,s} \overset{m}{\Gamma}_{jl}^{s} - \sum_{s=1}^{p} \overset{e}{\Gamma}_{ik,s} \overset{e}{\Gamma}_{jl}^{s}$$

$$+ \langle \overset{e}{A}_{jk} , \overset{m}{A}_{il} \rangle - \langle \overset{e}{A}_{jk} , \overset{e}{A}_{il} \rangle - g_{jk}g_{il} + \sum_{s=1}^{p} \overset{e}{\varGamma}_{jk,s} \overset{m}{\varGamma}_{il}^{s} - \sum_{s=1}^{p} \overset{e}{\varGamma}_{jk,s} \overset{e}{\varGamma}_{il}^{s}$$

$$+ \langle \overset{e}{A}_{ij} , \overset{m}{A}_{kl} \rangle - \langle \overset{e}{A}_{ij} , \overset{e}{A}_{kl} \rangle - g_{ij}g_{kl} + \sum_{s=1}^{p} \overset{e}{\varGamma}_{ij,s} \overset{m}{\varGamma}_{kl}^{s} - \sum_{s=1}^{p} \overset{e}{\varGamma}_{ij,s} \overset{e}{\varGamma}_{kl}^{s} \Big)$$

$$+ (\alpha' + 1)\partial_l \big(\overset{m}{\varGamma}_{ij,k} - \overset{e}{\varGamma}_{ij,k} \big) + \partial_l \overset{e}{\varGamma}_{ij,k} + \partial_l \overset{e}{\varGamma}_{ik,j} + \partial_l \overset{e}{\varGamma}_{kj,i}$$

$$+ \sum_{t=1}^{p} \big(\partial_k \overset{e}{\varGamma}_{ij}^{t} \big) g_{tl} + \sum_{t=1}^{p} \overset{e}{\varGamma}_{ij}^{t} \overset{e}{\varGamma}_{tk,l} - \langle \overset{e}{A}_{ij} , \overset{m}{A}_{kl} \rangle. \tag{93}$$

Acknowledgment

References

[1] S. Amari. Differential geometry of curved exponential families–curvature and information loss– *The Annals of Statistics*, 10:357-385, 1982.

[2] S. Amari. *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics 28. Springer-Verlag, 1985.

[3] S. Amari. Alpha divergence is unique, belonging to both classes of f-divergence and Bregman divergence *IEEE Trans. Information Theory*, 55:4925-4931, 2009.

[4] S. Amari and A. Chichocki. Information geometry of divergence function. *Bulletin of the Polish Academy of Sciences : Technical Sciences*, 58:183-195, 2010.

[5] S. Amari and H. Nagaoka. *Methods of Information Geometry*. Translations of Mathematical Monographs 191. American Mathematical Society, 2000.

[6] O. Calin and C. Udrişte. *Geometric modeling in probability and statistics*. Springer, 2014.

[7] J. Corcuera and F. Giummolè. On the relationship between alpha connections and the asymptotic properties of predictive distributions. *Bernoulli*, 5: 163-176, 1999.

[8] S. Eguchi. A differential geometric approach to statistical inference on the basis of contrast functionals. *Hiroshima Mathematical Journal*, 15: 341-391, 1985.

[9] S. Eguchi. Geometry of minimum contrast. *Hiroshima Mathematical Journal*, 22: 631-647, 1991.

[10] S. Eguchi and T. Yanagimoto Asymptotical improvement of maximum likelihood estimators on Kullback-Leibler loss. *Journal of Statistical Planning and Inference*, 138: 3502-3511, 2008.

[11] F. Komaki. On asymptotic properties of predictive distributions. *Biometrika*, 83: 299-313, 1996.

[12] A. Linde. A Bayesian view of model complexity. *Statistica Neerlandica*, 66: 253-271, 2012.

[13] P. Maji. F-information measures for efficient selection of discriminative genes from microarray data. *IEEE Transactions on Biomedical Engineering*, 56: 1063-1069, 2009

[14] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[15] M. K. Murray and J. W. Rice. *Differential Geometry and Statistics*. Chapman, 1993.

[16] Y.Qiao and N. Minematsu. A study on invariance of f-divergence and its application to speech recognition. *IEEE Transactions on Signal Processing*, 58: 3884-3890, 2010.

[17] Y. Sheena. Asymptotic expansion of the risk of maximum likelihood estimator with respect to α-divergence as a measure of the difficulty of specifying a parametric model –with detailed proof–. arXiv:1510.08226 [math.ST], 2017.

[18] L. T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11: 211-233, 1984.

[19] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkley Symposium on Mathematical and Statistical Probability*, 1: 197-206, University of California, Berkeley, 1956.

[20] K. Takeuchi and A. Takemura. Eigenfunctions of association algebra of pairings and zonal polynomials. *Discussion Paper 85-F-5, Faculty of Economics, University of Tokyo*, 1985.

[21] I. Vajda. *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, 1989.