Ann. Inst. Statist. Math.
Vol. 50, No. 3, 433–450 (1998)

A CHARACTERIZATION OF MONOTONE AND REGULAR DIVERGENCES

J. M. CORCUERA[1] AND F. GIUMMOLÈ[2]

[1] *Departament d'Estadística, Facultat de Matemàtiques, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain*
[2] *Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Via S. Francesco, 33, 35121 Padova, Italy*

(Received April 15, 1996; revised June 23, 1997)

Abstract. In this paper we characterize the local structure of monotone and regular divergences, which include f-divergences as a particular case, by giving their Taylor expansion up to fourth order. We extend a previous result obtained by Čencov, using the invariant properties of Amari's α-connections.

Key words and phrases: Differential geometry, divergence, embedding invariance, Markov embedding, α-connection.

1. Introduction

The necessity of measuring how different two populations are appears in many statistical problems. A wide class of indices or divergences has been used with such an objective (for a comprehensive exposition see Burbea (1983)). We are not able to give an universal rule for the choice of a divergence in each practical case. Anyway, we can investigate the general properties that an index of discrepancy should possess in order to describe a meaningful dissimilarity between populations. For instance, suppose to assemble the individuals of two finite populations in classes $A_1, \dots, A_m$. Let $D(P_1, P_2)$ be a convenient function of the proportions $P_i = (P_i(A_1), \dots, P_i(A_m))$, $i = 1, 2$, of individuals belonging to the different groups in the two populations. We can now decide to join several classes, obtaining $B_1, \dots, B_l$, $l < m$. If $\tilde{P}_i = (P_i(B_1), \dots, P_i(B_l))$, it is natural to demand that $D(\tilde{P}_1, \tilde{P}_2) \leq D(P_1, P_2)$, since the new classification brings less information than the previous one.

Divergences satisfying this property have already been studied by Čencov (1972). He gives their Taylor expansion up to second order, by means of the invariance of the Fisher metric. In this paper we extend Čencov's result to fourth order, using the invariance properties of the Amari α-connections and of a new class of fourth order tensors.

An additional property allows us to extend a divergence to the case when the individuals are classified in an infinite number of groups. This property expresses a sort of continuity of the divergence, when we let the number of classes tend to infinity.

2. Some basic definitions

In this section, we introduce operators representing an index of discrepancy between probability measures defined on the same measurable space.

In the sequel, we indicate with $(\mathcal{X}, \mathcal{A})$ a measurable space. $\mathcal{A}_\alpha \subseteq \mathcal{A}$ is a finite subσ-field of $\mathcal{A}$ and P_α is the restriction of P, defined on $(\mathcal{X}, \mathcal{A})$, to $\mathcal{A}_\alpha$.

DEFINITION 2.1. A *divergence* $D(P, Q)$ is a real-valued function whose arguments are two probability measures defined on the same measurable space.

DEFINITION 2.2. Let $(\mathcal{X}_1, \mathcal{A}_1)$ and $(\mathcal{X}_2, \mathcal{A}_2)$ be two measurable spaces. We say that $K : \mathcal{X}_1 \times \mathcal{A}_2 \to [0, 1]$ is a *Markov kernel*, $K \in Stoch\{(\mathcal{X}_1, \mathcal{A}_1), (\mathcal{X}_2, \mathcal{A}_2)\}$, if it satisfies the following properties:

1. $\forall A_2 \in \mathcal{A}_2$, $K(\cdot, A_2)$ is a measurable map;
2. $\forall x_1 \in \mathcal{X}_1$, $K(x_1, \cdot)$ is a probability on $(\mathcal{X}_2, \mathcal{A}_2)$.

If P is a probability measure on $(\mathcal{X}_1, \mathcal{A}_1)$, then K induces a probability measure on $(\mathcal{X}_2, \mathcal{A}_2)$, KP, defined by

$$KP(A_2) = \int_{\mathcal{X}_1} K(x, A_2) P(dx), \qquad \forall A_2 \subset \mathcal{A}_2.$$

Let $D(\cdot, \cdot)$ be a divergence and $(\mathcal{X}_1, \mathcal{A}_1)$ and $(\mathcal{X}_2, \mathcal{A}_2)$ be two measurable spaces.

DEFINITION 2.3. $D(P, Q)$ is said to be *monotone* with respect to Markov kernels if

$$-\infty < D(KP, KQ) \leq D(P, Q) \leq +\infty, \tag{2.1}$$

for every P, Q probability measures on $(\mathcal{X}_1, \mathcal{A}_1)$, and for every $K \in Stoch\{(\mathcal{X}_1, \mathcal{A}_1), (\mathcal{X}_2, \mathcal{A}_2)\}$.

As observed in the introduction, (2.1) is a natural property to require, since a transformation through a Markov kernel will, in general, cause a loss of information that is well explained by a decreasing of the divergence.

Monotonicity of a divergence function implies its invariance under a particular class of Markov kernels. Let $\mathcal{P}$ be a family of probabilities on $(\mathcal{X}_1, \mathcal{A}_1)$.

DEFINITION 2.4. $K \in Stoch\{(\mathcal{X}_1, \mathcal{A}_1), (\mathcal{X}_2, \mathcal{A}_2)\}$ is said to be *Blackwell sufficient (B-sufficient) with respect to* $\mathcal{P}$ if there exists $N \in Stoch\{(\mathcal{X}_2, \mathcal{A}_2), (\mathcal{X}_1, \mathcal{A}_1)\}$ such that $N(KP) = P$, $\forall P \in \mathcal{P}$. We say that K is B-sufficient if $\mathcal{P}$ is the family of all probability measures on $(\mathcal{X}_1, \mathcal{A}_1)$.

PROPOSITION 2.1. *If D is a monotone divergence with respect to Markov kernels, then, for every B-sufficient K,*

$$D(P,Q) = D(KP,KQ), \qquad \forall P,Q. \tag{2.2}$$

PROOF.

$$D(P,Q) = D(N(KP),N(KQ)) \leq D(KP,KQ),$$

that, together with the monotonicity, gives (2.2). □

This is also natural since a B-sufficient Markov kernel does not cause any loss of information.

COROLLARY 2.1. *The value of $D(P,P)$ is independent of P and it is a minimum value of the function D:*

$$D(P,P) = D_0 \leq D(Q,R), \qquad \forall P,Q,R.$$

PROOF. Given a probability measure P, there always exists a Markov kernel K, taking every probability measure to P: $K(x,\cdot) = P(\cdot)$, $\forall x \in \mathcal{X}$. Then,

$$D(Q,R) \geq D(KQ,KR) = D(P,P), \qquad \forall Q,R,$$

proving that $D(P,P)$ is a minimum value for D. Now, for every probability measure P', K is B-sufficient with respect to the family $\mathcal{P} = \{P,P'\}$, since $N(x,\cdot) = P'(\cdot)$, $\forall x \in \mathcal{X}$, transforms P into P'. By (2.2),

$$D(P,P) = D(P',P') = D_0, \qquad \forall P,P'.$$

□

DEFINITION 2.5. $D(P,Q)$ is said to be *regular* if

$$D(P,Q) = \lim_{\alpha} D(P_\alpha, Q_\alpha), \tag{2.3}$$

for every P and Q probability measures on $(\mathcal{X},\mathcal{A})$, where the limit is taken over the filter of all finite subσ-fields $\mathcal{A}_\alpha$ of $\mathcal{A}$, that is, over any increasing sequence $\{\mathcal{A}_n\}$ such that $\sigma(\bigcup_n \mathcal{A}_n) = \mathcal{A}$.

Remark. Since the restriction of a probability measure to a subσ-field is a particular case of Markov kernel, for monotone divergences the limit in (2.3) is a supremum.

The regularity condition enables us to extend to the general case a divergence originally defined on probability measures over finite σ-fields.

3. The multinomial case

Let $(\mathcal{X}, \mathcal{A}_m)$ be a measurable space, with $\mathcal{A}_m$ a finite subσ-field of $\mathcal{A}$ generated by the m atoms $A_1, \ldots, A_m$. Every probability measure P on $(\mathcal{X}, \mathcal{A})$ induces a probability measure on $(\mathcal{X}, \mathcal{A}_m)$, defined by m values, $x_1, \ldots, x_m$, with $x_i = P(A_i)$ and $\sum_{i=1}^m x_i = 1$. We shall consider only probability measures such that $x_i > 0$ for all $i = 1, \ldots, m$. Thus, for every P we have a point on the simplex

$$S_{m-1} = \left\{ x \in \mathbb{R}^m : \sum_{i=1}^m x_i = 1, x_i > 0, i = 1, \ldots, m \right\}.$$

S_{m-1} can be regarded as a surface in the differentiable manifold $\mathbb{R}^m$.

There is a tangent space T_x, with base $\{X_i = \frac{\partial}{\partial x_i}, i = 1, \ldots, m\}$, associated with every point $x \in \mathbb{R}^m$. If $x \in S_{m-1}$, the derivative of a function $h(x_1, \ldots, x_m)$ along a curve $x_i = \psi_i(t)$, $i = 1, \ldots, m$, tangent to S_{m-1}, takes the form $\sum_{i=1}^m \psi_i'(t) \frac{\partial h}{\partial x_i}$. Since $\sum_{i=1}^m \psi_i(t) = 1$, then $\sum_{i=1}^m \psi_i'(t) = 0$ and every vector X tangent to the simplex S_{m-1} can be represented as $X = \sum_{i=1}^m a_i X_i$, with $\sum_{i=1}^m a_i = 0$. Let us denote by M_x the tangent space to S_{m-1} in x. It is easy to see that $U_i = X_i - X_m$, $i = 1, \ldots, m-1$, belong to M_x and are independent. We can thus take $\{U_i = \frac{\partial}{\partial x_i} - \frac{\partial}{\partial x_m}, i = 1, \ldots, m-1\}$ as a base of M_x, for every $x \in S_{m-1}$.

For $n \geq m$, let $B_1, \ldots, B_n$ be a partition of $\mathcal{X}$ such that $A_i = \bigcup_{j \in I_i} B_j$, $i = 1, \ldots, m$, where $I_1, \ldots, I_m$ is a partition of $\{1, \ldots, n\}$. For any probability measure P on $(\mathcal{X}, \mathcal{A})$, let $x_i = P(A_i)$, $i = 1, \ldots, m$, and $y_j = P(B_j)$, $j = 1, \ldots, n$. Thus,

$$x_i = \sum_{j \in I_i} y_j, \qquad i = 1, \ldots, m. \tag{3.1}$$

Conversely, define

$$q_{ij} = P(B_j \mid A_i) = \begin{cases} y_j / x_i & \text{if } j \in I_i \\ 0 & \text{if } j \notin I_i. \end{cases}$$

Then,

$$y_j = \sum_{i=1}^m q_{ij} x_i, \qquad j = 1, \ldots, n. \tag{3.2}$$

Notice that (q_{ij}) is a stochastic matrix, that is: $q_{ij} \geq 0$, $\forall i, j$ and $\sum_{j=1}^n q_{ij} = 1$, $\forall i$. Moreover, for every j, $q_{rj} q_{sj} = 0$ if $r \neq s$, and there is some i such that $q_{ij} > 0$. Then, for every stochastic matrix with these properties, (3.2) defines a function $f : \mathbb{R}^m \to \mathbb{R}^n$ which inverse is given by (3.1). The restriction of f to S_{m-1} is a B-sufficient Markov kernel, with respect to the family of all probability measures on $(\mathcal{X}, \mathcal{A}_m)$. In fact, it is easy to prove that any B-sufficient Markov kernel can be written in the form (3.2). We call f a *Markov embedding*.

The Jacobian map associated with f, $f^* : T_x \to T_y$, is defined by

$$f^* X_i = \sum_{j=1}^{n} q_{ij} Y_j, \qquad i = 1, \dots, m, \tag{3.3}$$

with $Y_j = \frac{\partial}{\partial y_j}$, $j = 1, \dots, n$.

3.1 *Embedding invariant structures*

In the present section we consider geometrical structures defined on simplexes, that are invariant with respect to Markov embeddings. We characterize invariant Riemannian metrics and affine connections, showing that, up to constant factors, they coincide respectively with the Fisher metric and the Amari α-connections.

3.1.1 *Embedding invariant divergences*

From now on, we consider in S_{m-1} the system of local coordinates that with each P on $(\mathcal{X}, \mathcal{A})$ associates a vector $x = (x_1, \dots, x_{m-1})$, where $x_i = P(A_i)$, $i = 1, \dots, m$, and $\sum_{i=1}^{m} x_i = 1$. Every Markov embedding $f : S_{m-1} \to S_{n-1}$ can be expressed in local coordinates as

$$y_j = \sum_{i=1}^{m-1} (q_{ij} - q_{mj}) x_i + q_{mj} = \sum_{i=1}^{m-1} \bar{q}_{ij} x_i + q_{mj}, \qquad j = 1, \dots, n-1 \tag{3.4}$$

and the Jacobian map $f^* : M_x \to M_y$ associated with f, as

$$f^* U_i = \sum_{j=1}^{n-1} (q_{ij} - q_{mj}) V_j = \sum_{j=1}^{n-1} \bar{q}_{ij} V_j, \qquad i = 1, \dots, m-1, \tag{3.5}$$

where $\{V_j = \frac{\partial}{\partial y_j} - \frac{\partial}{\partial y_n}, j = 1, \dots, n-1\}$ is the base of M_y in S_{n-1}.

Let D be a divergence, P and Q equivalent probability measures on $(\mathcal{X}, \mathcal{A})$, P_m and Q_m respectively the induced probabilities in $(\mathcal{X}, \mathcal{A}_m)$ and x and $\tilde{x}$ the corresponding expressions in local coordinates in S_{m-1}. Thus,

$$\overset{m}{d}(x, \tilde{x}) = D(P_m, Q_m)$$

is the expression of D in local coordinates in S_{m-1}, so that, as m changes, the divergence D induces a whole family $\overset{m}{d}$ of real valued functions defined on the simplexes.

DEFINITION 3.1. $D(P, Q)$ is said to be *embedding invariant* (*e-invariant*) if it is invariant with respect to every Markov embedding $f : S_{m-1} \to S_{n-1}$, for each m and $n \geq m$.

The e-invariance property of a divergence can be expressed in local coordinates as

$$\overset{m}{d}(x, \tilde{x}) = \overset{n}{d}(y, \tilde{y}), \qquad \forall x, \tilde{x} \in S_{m-1}, \tag{3.6}$$

where the components of y and $\tilde{y}$ are respectively related to those of x and $\tilde{x}$ by (3.4).

It is important to notice that, since every Markov embedding is B-sufficient, by Proposition 2.1, every monotone divergence satisfies the preceding condition, being thus e-invariant.

3.1.2 *Invariant Riemannian metrics*

Let $\langle \cdot, \cdot \rangle$ denote the family of scalar products $\langle \cdot, \cdot \rangle_m(x)$ on the tangent space M_x, $x \in S_{m-1}$, for different values of m and x.

DEFINITION 3.2. $\langle \cdot, \cdot \rangle$ is said to be *e-invariant* if

$$\langle U, \tilde{U} \rangle_m(x) = \langle f^*U, f^*\tilde{U} \rangle_n(y), \qquad \forall U, \tilde{U} \in M_x, \quad \forall x \in S_{m-1}, \tag{3.7}$$

for every Markov embedding $f : S_{m-1} \to S_{n-1}$ and for each m and $n \geq m$.

Let $\overset{m}{g}_{ij}(x) = \langle U_i, U_j \rangle_m(x)$ and $\overset{n}{g}_{rs}(y) = \langle V_r, V_s \rangle_n(y)$ denote respectively the components of the metric tensor in S_{m-1} and in S_{n-1}. If f is a Markov embedding between S_{m-1} and S_{n-1}, then $\langle \cdot, \cdot \rangle_n(y)$ naturally induces a scalar product on its $m-1$-dimensional submanifold $f(S_{m-1})$:

$$\overset{n}{\bar{g}}_{ij}(x) = \langle f^*U_i, f^*U_j \rangle_n(y), \qquad i, j = 1, \ldots, m-1.$$

We can then rewrite condition (3.7) as

$$\overset{m}{g}_{ij}(x) = \overset{n}{\bar{g}}_{ij}(y), \qquad i, j = 1, \ldots, m-1 \tag{3.8}$$

and, using (3.5), the preceding definition can be written equivalently as

$$\overset{m}{g}_{ij}(x) = \sum_{r,s=1}^{n-1} \overset{n}{g}_{rs}(y) \bar{q}_{ir} \bar{q}_{js}, \qquad i, j = 1, \ldots, m-1. \tag{3.9}$$

The following result was first given by Čencov (1972), anyway, we refer the reader to Campbell (1986) for an easier proof. It essentially states the unicity of the Fisher metric as an e-invariant Riemannian metric, see also Amari ((1985), p. 31).

THEOREM 3.1. *The only e-invariant Riemannian metrics are of the form*

$$\overset{m}{g}_{ij}(x) = \langle U_i, U_j \rangle_m(x) = A \left(\frac{\delta_{ij}}{x_i} + \frac{1}{x_m} \right), \qquad i, j = 1, \ldots, m-1, \tag{3.10}$$

where $A > 0$ and δ_{ij} is the Kronecker delta.

3.1.3 *Invariant affine connections*

A characterization similar to that given for e-invariant metrics, can be obtained for affine connections.

Let ∇ denote the family of affine connections $\overset{m}{\nabla}$ defined on S_{m-1} for different values of m.

If f is a Markov embedding between S_{m-1} and S_{n-1}, $m \leq n$, then $\overset{m}{\nabla}$ induces through f an affine connection on $f(S_{m-1})$, defined by

$$f^*\left(\overset{m}{\nabla}_U \tilde{U}(x)\right), \qquad \forall x \in S_{m-1}$$

where U, $\tilde{U}$ are arbitrary smooth vector fields on S_{m-1}. Moreover, as a submanifold of the Riemannian manifold S_{n-1}, $f(S_{m-1})$ naturally inherits the affine connection of S_{n-1}:

$$\overset{n}{\bar{\nabla}}_{f^*U}\, f^*\tilde{U}(y),$$

where $\overset{n}{\bar{\nabla}}$ is the orthogonal projection of $\overset{n}{\nabla}$ on the tangent space to $f(S_{m-1})$.

DEFINITION 3.3. ∇ is said to be *e-invariant* if the affine connection induced by $\overset{m}{\nabla}$ on $f(S_{m-1})$ through f coincides with that induced on $f(S_{m-1})$ by $\overset{n}{\nabla}$, that is,

$$f^*\left(\overset{m}{\nabla}_U \tilde{U}(x)\right) = \overset{n}{\bar{\nabla}}_{f^*U}\, f^*\tilde{U}(y), \qquad \forall U, \tilde{U}, \qquad \forall x \in S_{m-1}, \tag{3.11}$$

for every Markov embedding $f : S_{m-1} \to S_{n-1}$ and for each m and $n \geq m$.

Remark. It is important to observe that condition (3.11) can be expressed in the equivalent form

$$\overset{m}{\Gamma}_{ijk}(x) = \sum_{r,s,t=1}^{n-1} \overset{n}{\Gamma}_{rst}(y)\bar{q}_{ir}\bar{q}_{js}\bar{q}_{kt}, \qquad i,j,k = 1,\ldots,m-1, \tag{3.12}$$

where $\overset{m}{\Gamma}$ and $\overset{n}{\Gamma}$ denote respectively the coefficients of the affine connection in S_{m-1} and S_{n-1}, with respect to the usual bases of M_x and M_y. In fact, let g be any e-invariant metric, so that (3.8) is satisfied. We thus have, using the repeated index convention,

$$f^*\left(\overset{m}{\nabla}_{U_i} U_j\right) = \overset{m}{\Gamma}_{ij}{}^h f^* U_h = \sum_{u=1}^{n-1} \overset{m}{\Gamma}_{ij}{}^h \bar{q}_{hu} V_u = \sum_{u=1}^{n-1}\left(\overset{m}{\Gamma}_{ijk} \overset{m}{g}{}^{kh} \bar{q}_{hu}\right) V_u$$

and

$$
\begin{aligned}
\overset{n}{\bar{\nabla}}_{f^*U_i} f^*U_j &= \sum_{r,s=1}^{n-1} \overset{n}{\Gamma}_{rs}{}^{w} \bar{q}_{ir}\bar{q}_{js} \langle V_w, f^*U_k\rangle_n \overset{n}{\bar{g}}{}^{kh} f^*U_h \\
&= \sum_{r,s,t,u=1}^{n-1} \overset{n}{\Gamma}_{rs}{}^{w} \bar{q}_{ir}\bar{q}_{js}\bar{q}_{kt}\bar{q}_{hu} \overset{n}{g}_{wt} \overset{n}{\bar{g}}{}^{kh} V_u \\
&= \sum_{u=1}^{n-1} \left(\sum_{r,s,t=1}^{n-1} \overset{n}{\Gamma}_{rst} \bar{q}_{ir}\bar{q}_{js}\bar{q}_{kt}\bar{q}_{hu} \overset{m}{g}{}^{kh} \right) V_u,
\end{aligned}
$$

that immediately give (3.12).

The following theorem gives a characterization for e-invariant affine connections, showing that they coincide, up to a constant factor, with the Amari α-connections in the multinomial case (Amari (1985), p. 43).

THEOREM 3.2. *The only e-invariant affine connections have coefficients of the form*

$$
\overset{m}{\Gamma}_{ijk}(x) = B\left(\frac{1}{x_m^2} - \frac{\delta_{ijk}}{x_i^2}\right), \qquad i,j,k = 1,\ldots,m-1, \tag{3.13}
$$

where B is a constant.

The proof is given in the Appendix.

Remark. (3.9) and (3.12) suggest an obvious way to extend the definition of e-invariance to every array of order k such that

$$
\overset{m}{T}_{i_1,\ldots,i_k}(x) = \sum_{r_1,\ldots,r_k=1}^{n-1} \overset{n}{T}_{r_1,\ldots,r_k}(y)\bar{q}_{i_1r_1},\ldots,\bar{q}_{i_kr_k}. \tag{3.14}
$$

It is readily seen that (3.10) with $A \in \mathbb{R}$ and (3.13) provide respectively a characterization for second and third order e-invariant arrays.

3.2 *The local structure of D*

To study the local behaviour of a divergence D, suppose it is smooth, that is, at any point of S_{m-1} it admits an expression in local coordinates that is differentiable up to necessary order.

3.2.1 *The geometry of monotone divergences*

Fixed m, every monotone divergence induces on S_{m-1} a collection of geometric objects. We can express the condition of minimum in the diagonal by:

$$
\begin{aligned}
&\overset{m}{d}(x,x) = \overset{m}{d}_0, \\
&\overset{m}{d}_{;i}(x,x) = 0
\end{aligned} \tag{3.15}
$$

and

$$\overset{m}{d}_{;ij}(x,x) \geq 0, \tag{3.16}$$

for every $x \in S_{m-1}$, where the semicolon indicates the argument with respect to which the derivative is taken.

By (3.15) and (3.16), we have a second order covariant tensor, $\overset{m}{d}_{;ij}$, associated with $\overset{m}{d}$. When strictly positive definite, $\overset{m}{d}_{;ij}$ is a metric tensor. It is important to notice that in fact it is a second order e-invariant array, since, by derivating twice (3.6),

$$\overset{m}{d}_{;ij}(x,x) = \sum_{r,s=1}^{n-1} \overset{n}{d}_{;rs}(y,y)\bar{q}_{ir}\bar{q}_{js},$$

for every $i, j = 1, \ldots, m-1$.

$\overset{m}{d}$ also generates an affine connection with coefficients

$$\overset{m}{\Gamma}_{ijk}(x) = -\overset{m}{d}_{k;ij}(x,x).$$

In fact, it is an e-invariant affine connection, since

$$\overset{m}{d}_{k;ij}(x,x) = \sum_{r,s,t=1}^{n-1} \overset{n}{d}_{t;rs}(y,y)\bar{q}_{ir}\bar{q}_{js}\bar{q}_{kt},$$

for every $i, j, k = 1, \ldots, m-1$. Actually, it is easily seen that the partial derivatives of any order of an e-invariant divergence, satisfy (3.14). They could be called an e-invariant string (Blæsild (1991)). This observation, together with the results of the previous section, allows us to characterize the Taylor expansion of any monotone divergence in the simplex.

THEOREM 3.3. *For each P and Q in $(\mathcal{X}, \mathcal{A})$ and any monotone divergence $D(P,Q)$, we have the expansion*

$$\begin{aligned} D(P_m, Q_m) = D_0 &+ D_1 \sum_{i=1}^{m} \frac{[Q(A_i) - P(A_i)]^2}{P(A_i)} + D_2 \sum_{i=1}^{m} \frac{[Q(A_i) - P(A_i)]^3}{P(A_i)^2} \\ &+ D_3 \sum_{i=1}^{m} \frac{[Q(A_i) - P(A_i)]^4}{P(A_i)^3} + D_4 \left(\sum_{i=1}^{m} \frac{[Q(A_i) - P(A_i)]^2}{P(A_i)} \right)^2 \\ &+ o(\|Q_m - P_m\|^4), \end{aligned}$$

where P_m and Q_m are the induced probabilities on $(\mathcal{X}, \mathcal{A}_m)$ and D_0, D_1, D_2, D_3, D_4 are constants, $D_1 \geq 0$.

PROOF. Let $x = (x_1, \ldots, x_{m-1})$ and $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_{m-1})$ be such that $x_i = P(A_i)$, $\tilde{x}_i = Q(A_i)$, $i = 1, \ldots, m$, $\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} \tilde{x}_i = 1$. Then, writing $d(x, \tilde{x}) =$

$D(P_m, Q_m)$ and $(\tilde{x}-x)_I = (\tilde{x}_{i_1} - x_{i_1}) \cdots (\tilde{x}_{i_k} - x_{i_k})$, with $I = (i_1, \ldots, i_k)$, we have

$$d(x,\tilde{x}) = d(x,x) + \sum_{i=1}^{m-1} d_{;i}(x,x)(\tilde{x}-x)_i + \frac{1}{2}\sum_{i,j=1}^{m-1} d_{;ij}(x,x)(\tilde{x}-x)_{ij} + \frac{1}{6}\sum_{i,j,k=1}^{m-1} d_{;ijk}(x,x)(\tilde{x}-x)_{ijk} + \frac{1}{24}\sum_{i,j,k,l=1}^{m-1} d_{;ijkl}(x,x)(\tilde{x}-x)_{ijkl} + o(\|\tilde{x}-x\|^4).$$

By the monotonicity of D, (3.10) and (3.13), we obtain

$$\sum_{i,j=1}^{m-1} d_{;ij}(x,x)(\tilde{x}-x)_{ij} = A\sum_{i,j=1}^{m-1}\left(\frac{\delta_{ij}}{x_i} + \frac{1}{x_m}\right)(\tilde{x}-x)_{ij} = A\left(\sum_{i=1}^{m-1}\frac{(\tilde{x}-x)_i^2}{x_i} + \sum_{i,j=1}^{m-1}\frac{(\tilde{x}-x)_{ij}}{x_m}\right) = A\left(\sum_{i=1}^{m-1}\frac{(\tilde{x}-x)_i^2}{x_i} + \frac{(\tilde{x}-x)_m^2}{x_m}\right) = A\sum_{i=1}^{m}\frac{(\tilde{x}-x)_i^2}{x_i}, \tag{3.17}$$

since $(\tilde{x}-x)_m = -\sum_{i=1}^{m-1}(\tilde{x}-x)_i$, and

$$\sum_{i,j,k=1}^{m-1} d_{;ijk}(x,x)(\tilde{x}-x)_{ijk} = B\sum_{i,j,k=1}^{m-1}\left(\frac{1}{x_m^2} - \frac{\delta_{ijk}}{x_i^2}\right)(\tilde{x}-x)_{ijk} = -B\sum_{i=1}^{m}\frac{(\tilde{x}-x)_i^3}{x_i^2}. \tag{3.18}$$

As regards the fourth order term, it can be shown that any fourth order e-invariant array in S_{m-1} has the form

$$T_{ijkl}(x) = C\left\{\left(\frac{1}{x_m}+\frac{\delta_{il}}{x_i}\right)\left(\frac{1}{x_m}+\frac{\delta_{jk}}{x_j}\right) + \left(\frac{1}{x_m}+\frac{\delta_{jl}}{x_j}\right)\left(\frac{1}{x_m}+\frac{\delta_{ik}}{x_k}\right) + \left(\frac{1}{x_m}+\frac{\delta_{kl}}{x_k}\right)\left(\frac{1}{x_m}+\frac{\delta_{ij}}{x_i}\right)\right\} + D\left(\frac{1}{x_m^3}+\frac{\delta_{ijkl}}{x_i^3}\right)$$

(see the Appendix for the proof). Thus,

$$\sum_{i,j,k,l=1}^{m-1} d_{;ijkl}(x,x)(\tilde{x}-x)_{ijkl} \tag{3.19}$$

$$= 3C\left[\sum_{j,k=1}^{m-1}\left(\frac{1}{x_m}+\frac{\delta_{jk}}{x_j}\right)(\tilde{x}-x)_{jk}\sum_{i,l=1}^{m-1}\left(\frac{1}{x_m}+\frac{\delta_{il}}{x_i}\right)(\tilde{x}-x)_{il}\right]$$
$$+ D\sum_{i=1}^{m}\frac{(\tilde{x}-x)_i^4}{x_i^3}$$
$$= 3C\left(\sum_{i=1}^{m}\frac{(\tilde{x}-x)_i^2}{x_i}\right)^2 + D\sum_{i=1}^{m}\frac{(\tilde{x}-x)_i^4}{x_i^3}$$

and the proposition follows by (3.17), (3.18) and (3.19). □

4. The general case

In the preceding section, we obtained a local expression for monotone divergences defined on multinomial distributions. Using the regularity condition, we are able to extend this expansion to the general case. We need the following result:

THEOREM 4.1. *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex function. For any P and Q, equivalent probability measures on $(\mathcal{X}, \mathcal{A})$, there exists a non negative integral*

$$\int_{\mathcal{X}} f\left(\frac{dQ}{dP}(x)\right)P(dx) = \lim_{\alpha}\sum_i f\left(\frac{Q(A_i)}{P(A_i)}\right)P(A_i),$$

where the limit is taken over the filter of all finite subσ-fields $\mathcal{A}_\alpha$ of $\mathcal{A}$.

PROOF. Since

$$\sum_i f\left(\frac{Q(A_i)}{P(A_i)}\right)P(A_i) = \int_{\mathcal{X}} f\left(\frac{dQ_\alpha}{dP_\alpha}(x)\right)P_\alpha(dx),$$

the thesis can be written in the form

$$\lim_{\alpha}\int_{\mathcal{X}} f\left(\frac{dQ_\alpha}{dP_\alpha}(x)\right)P_\alpha(dx) = \int_{\mathcal{X}} f\left(\frac{dQ}{dP}(x)\right)P(dx).$$

It is sufficient to prove it for any increasing sequence $\{\mathcal{A}_n\}$ of finite subσ-field of $\mathcal{A}$, such that $\sigma(\bigcup_n \mathcal{A}_n) = \mathcal{A}$. Since f-divergences are monotone, see Heyer ((1982), Theorem 22.9, p. 169), and by the remark following Definition 2.5, we obtain:

$$\lim_{n}\int_{\mathcal{X}} f\left(\frac{dQ_n}{dP_n}(x)\right)P_n(dx) \leq \int_{\mathcal{X}} f\left(\frac{dQ}{dP}(x)\right)P(dx).$$

We show now that the reverse inequality also holds. Since

$$\int_{\mathcal{X}} \frac{dQ}{dP}(x)P(dx) = 1 < \infty,$$

we can apply a well known theorem of convergence of martingales, thus obtaining

$$\frac{dQ_n}{dP_n} = E\left(\frac{dQ}{dP} \,\middle|\, a_n\right) \overset{\text{a.s.}}{\to} E\left(\frac{dQ}{dP} \,\middle|\, a\right) = \frac{dQ}{dP}.$$

Since f is a continuous function, the convergence still holds:

$$f\left(\frac{dQ_\alpha}{dP_\alpha}\right) \overset{\text{a.s.}}{\to} f\left(\frac{dQ}{dP}\right).$$

By the Fatou lemma,

$$\begin{aligned}\int_{\mathcal{X}} f\left(\frac{dQ}{dP}(x)\right) P(dx) &\leq \lim_n \int_{\mathcal{X}} f\left(\frac{dQ_n}{dP_n}(x)\right) P(dx) \\ &= \lim_n \int_{\mathcal{X}} f\left(\frac{dQ_n}{dP_n}(x)\right) P_n(dx),\end{aligned}$$

and the thesis is proved. □

Remark. The preceding theorem can be easily extended to the case of f being any linear combination of non negative convex functions.

We can now prove the main result of the present section:

THEOREM 4.2. *If*

$$\int_{\mathcal{X}} \left|\frac{Q(dx) - P(dx)}{P(dx)}\right|^4 P(dx) < \infty,$$

then, at each point P, *any monotone and regular divergence* $D(P,Q)$, *admits the expansion*

$$\begin{aligned}(4.1)\quad D(P,Q) = D_0 &+ D_1 \int_{\mathcal{X}} \frac{[Q(dx) - P(dx)]^2}{P(dx)} + D_2 \int_{\mathcal{X}} \frac{[Q(dx) - P(dx)]^3}{P(dx)^2} \\ &+ D_3 \int_{\mathcal{X}} \frac{[Q(dx) - P(dx)]^4}{P(dx)^3} + D_4 \left(\int_{\mathcal{X}} \frac{[Q(dx) - P(dx)]^2}{P(dx)}\right)^2 \\ &+ o(\|Q-P\|^4),\end{aligned}$$

where $\|Q-P\|^4 = \int_{\mathcal{X}} \frac{[Q(dx)-P(dx)]^4}{P(dx)^3}$ *and* D_0, D_1, D_2, D_3, D_4 *are constants,* $D_1 \geq 0$.

PROOF. By Theorem 3.3, it holds:

$$\begin{aligned}D(P_\alpha, Q_\alpha) = D_0 &+ D_1 \sum_i \frac{[Q(A_i) - P(A_i)]^2}{P(A_i)} + D_2 \sum_i \frac{[Q(A_i) - P(A_i)]^3}{P(A_i)^2} \\ &+ D_3 \sum_i \frac{[Q(A_i) - P(A_i)]^4}{P(A_i)^3} + D_4 \left(\sum_i \frac{[Q(A_i) - P(A_i)]^2}{P(A_i)}\right)^2 \\ &+ o(\|Q_\alpha - P_\alpha\|^4),\end{aligned}$$

for every P_α and Q_α, restrictions of P and Q to the finite dimensional subσ-field $\mathcal{A}_\alpha$ of $\mathcal{A}$. We can now pass to the limit. The terms with coefficients D_1 and D_4 can be obtained by applying Theorem 4.1 with $f(x) = (x-1)^2$. The same holds for the term with coefficient D_3, with $f(x) = (x-1)^4$. For the third order term we can use the remark following Theorem 4.1, since $f(x) = (x-1)^3$ can be written as the difference of two non negative convex functions:

$$f(x) = f_1(x) - f_2(x),$$

where

$$f_1(x) = \begin{cases} 0 & x \le 1 \\ (x-1)^3 & x > 1 \end{cases}$$

and

$$f_2(x) = \begin{cases} -(x-1)^3 & x \le 1 \\ 0 & x > 1. \end{cases}$$

The regularity of D guarantees the result. □

4.1 *The parametric case*

Suppose now that P and Q belong to some regular parametric model, that is, P and Q are equivalent probability measures with densities $p(x;\theta)$ and $p(x;\theta')$, $\theta, \theta' \in \Theta \subset \mathbb{R}^k$, with respect to some common dominating measure μ. By Theorem 4.2, we have that any monotone and regular divergence between P and Q can be expanded as:

$$\begin{aligned} (4.2)\quad D(P,Q) &= D(\theta,\theta') \\ &= D_0 + D_1 \int \left(\frac{p(x;\theta') - p(x;\theta)}{p(x;\theta)}\right)^2 p(x;\theta)\mu(dx) \\ &+ D_2 \int \left(\frac{p(x;\theta') - p(x;\theta)}{p(x;\theta)}\right)^3 p(x;\theta)\mu(dx) \\ &+ D_3 \int \left(\frac{p(x;\theta') - p(x;\theta)}{p(x;\theta)}\right)^4 p(x;\theta)\mu(dx) \\ &+ D_4 \left[\int \left(\frac{p(x;\theta') - p(x;\theta)}{p(x;\theta)}\right)^2 p(x;\theta)\mu(dx)\right]^2 + o(|\theta' - \theta|^4). \end{aligned}$$

Moreover, writing $(\theta' - \theta)_{I_r} = (\theta'_{i_1} - \theta_{i_1}) \cdots (\theta'_{i_r} - \theta_{i_r})$ and $\partial_{I_r} = \partial^r / \partial\theta_{i_1} \cdots \partial\theta_{i_r}$, with $I_r = (i_1, \ldots, i_r)$, and using the repeated index convention, we have

$$\begin{aligned} p(x;\theta') &= p(x;\theta) + \partial_i p(x;\theta)(\theta' - \theta)_i + \frac{1}{2}\partial_{ij} p(x;\theta)(\theta' - \theta)_{ij} \\ &+ \frac{1}{6}\partial_{ijk} p(x;\theta)(\theta' - \theta)_{ijk} + o(|\theta' - \theta|^3), \end{aligned}$$

so that:

$$\left(\frac{p(x;\theta')-p(x;\theta)}{p(x;\theta)}\right)^2$$
$$= \left(\frac{\partial_i p(x;\theta)}{p(x;\theta)}(\theta'-\theta)_i + \frac{1}{2}\frac{\partial_{ij}p(x;\theta)}{p(x;\theta)}(\theta'-\theta)_{ij} + \frac{1}{6}\frac{\partial_{ijk}p(x;\theta)}{p(x;\theta)}(\theta'-\theta)_{ijk}\right)^2$$
$$+ o(|\theta'-\theta|^4)$$
$$= \partial_i l \partial_j l(\theta'-\theta)_{ij} + \partial_k l\,(\partial_{ij}l + \partial_i l\partial_j l)\,(\theta'-\theta)_{ijk}$$
$$+ \left(\frac{1}{3}\partial_{ijk}l\partial_h l + \frac{1}{4}\partial_{ij}l\partial_{kh}l + \frac{3}{2}\partial_{ij}l\partial_k l\partial_h l + \frac{7}{12}\partial_i l\partial_j l\partial_k l\partial_h l\right)(\theta'-\theta)_{ijkh}$$
$$+ o(|\theta'-\theta|^4),$$
$$\left(\frac{p(x;\theta')-p(x;\theta)}{p(x;\theta)}\right)^3$$
$$= \left(\frac{\partial_i p(x;\theta)}{p(x;\theta)}(\theta'-\theta)_i + \frac{1}{2}\frac{\partial_{ij}p(x;\theta)}{p(x;\theta)}(\theta'-\theta)_{ij}\right)^3 + o(|\theta'-\theta|^4)$$
$$= \partial_i l\partial_j l\partial_k l(\theta'-\theta)_{ijk} + \frac{3}{2}\,(\partial_{ij}l\partial_k l\partial_h l + \partial_i l\partial_j l\partial_k l\partial_h l)\,(\theta'-\theta)_{ijkh}$$
$$+ o(|\theta'-\theta|^4)$$

and

$$\left(\frac{p(x;\theta')-p(x;\theta)}{p(x;\theta)}\right)^4 = \partial_i l\partial_j l\partial_k l\partial_h l(\theta'-\theta)_{ijkh} + o(|\theta'-\theta|^4).$$

By substituting in (4.2) and writing $\nu_{I_1,\dots,I_r}(\theta) = E_\theta(\partial_{I_1}l\cdots\partial_{I_r}l)$, for some multiindices $I_1,\dots,I_r$, we obtain

$$\begin{aligned}
D(\theta,\theta') &= D_0 + D_1\nu_{i,j}(\theta)(\theta'-\theta)_{ij} + [D_1\nu_{ij,k}(\theta) + (D_1+D_2)\nu_{i,j,k}(\theta)](\theta'-\theta)_{ijk} \\
&\quad + \left[\frac{D_1}{3}\nu_{ijk,h}(\theta) + \frac{D_1}{4}\nu_{ij,kh}(\theta) + \frac{3}{2}(D_1+D_2)\nu_{ij,k,h}(\theta)\right. \\
&\qquad \left. + \left(\frac{7}{12}D_1 + \frac{3}{2}D_2 + D_3\right)\nu_{i,j,k,h}(\theta) + D_4\nu_{i,j}(\theta)\nu_{k,h}(\theta)\right](\theta'-\theta)_{ijkh} \\
&\quad + o(|\theta'-\theta|^4) \\
&= D_0 + D_1\left[g_{ij}(\theta)(\theta'-\theta)_{ij} + \overset{\alpha}{\Gamma}_{ijk}(\theta)(\theta'-\theta)_{ijk}\right] \\
&\quad + \left[\frac{D_1}{3}\nu_{ijk,h}(\theta) + \frac{D_1}{4}\nu_{ij,kh}(\theta) + \frac{3}{2}(D_1+D_2)\nu_{ij,k,h}(\theta)\right. \\
&\qquad \left. + \left(\frac{7}{12}D_1 + \frac{3}{2}D_2 + D_3\right)\nu_{i,j,k,h}(\theta) + D_4 g_{ij}(\theta)g_{kh}(\theta)\right](\theta'-\theta)_{ijkh} \\
&\quad + o(|\theta'-\theta|^4),
\end{aligned}$$

where $\alpha = -\frac{D_1+2D_2}{D_1}$, $g_{ij}(\theta)$ and $\overset{\alpha}{\Gamma}_{ijk}(\theta)$ are respectively the Fisher metric and Amari's α-connections of the parametric model, see Amari ((1985), pp. 26 and 39).

4.2 *Concluding remarks*

Monotonicity and regularity are the properties that every measure of discrepancy between probabilities should possess. In this sense, we believe that all meaningful inference procedures should be deduced from some monotone and regular divergence. The characterization given here can thus be useful for studying the properties of such procedures.

It is also interesting to point out that the contrast functions used by Eguchi (1992) are measures of discrepancy between probabilities belonging to a certain family. Then, since the result of applying a Markov kernel is, in general, a probability out of the family, the requirement of monotonicity becomes meaningless.

5. Appendix

PROOF OF THEOREM 3.2. In order to prove the theorem, we show that the linear space generated by all the e-invariant affine connections on the simplexes, has dimension one. (3.13) then follows from the fact that a third order e-invariant array can be obtained by derivating a second order array of the form (3.10).

First of all, we show that the Christoffel symbols of an e-invariant affine connection, calculated in the center of the simplex S_{m-1}, depend only on a constant. Consider the case of $m = n$ and let f be the Markov embedding that interchanges two of the first $m-1$ coordinates. Then, $f(x) = x$ for $x = (1/m, \ldots, 1/m)$. By (3.12), we can easily see that $\overset{m}{\Gamma}$ calculated in x is completely symmetric and:

$$\overset{m}{\Gamma}_{iii} = F_m, \qquad \overset{m}{\Gamma}_{iik} = G_m, \qquad \overset{m}{\Gamma}_{ijk} = H_m,$$

where different indices are suppose to take different values. If f interchanges some coordinate x_j and x_m, we have that, in the center x,

$$\overset{m}{\Gamma}_{jjj} = 0,$$

and

$$\overset{m}{\Gamma}_{ijk} = -\overset{m}{\Gamma}_{ijk} + \overset{m}{\Gamma}_{ijj} - \overset{m}{\Gamma}_{jjj} + \overset{m}{\Gamma}_{jji},$$

that is,

$$F_m = 0 \quad \text{and} \quad G_m = H_m.$$

Thus, for $x = (1/m, \ldots, 1/m)$, we can write

$$\overset{m}{\Gamma}_{ijk}(x) = \begin{cases} 0 & i = j = k \\ G_m & \text{otherwise,} \end{cases}$$

and the value of $\overset{m}{\Gamma}$ in the center of S_{m-1} depends only on a constant.

Notice now that the value of G_m for a fixed $m > 3$, determines G_n for each $n \neq m$. In fact, let $n = hm$, with h an integer bigger than one and consider the Markov embedding defined by $f_h(x) = (x_1/h, \ldots, x_1/h, \ldots, x_m/h, \ldots, x_m/h)$, each component being repeated h times. Thus f_h maps $x = (1/m, \ldots, 1/m)$ to $y = (1/n, \ldots, 1/n)$ and

$$\bar{q}_{ia} = \begin{cases} 1/h & a \in R_i \\ -1/h & a \in R_m \setminus \{n\} \\ 0 & \text{otherwise}, \end{cases} \tag{5.1}$$

where $R_i = \{(i-1)h+1, \ldots, ih\}$, $i = 1, \ldots, m$. By (3.12), we have that

$$\overset{m}{\Gamma}_{ijk} = \overset{n}{\Gamma}_{aaa} \sum_{a=1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{ka} + 3\, \overset{n}{\Gamma}_{aab} \sum_{a,b-1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{kb} + \overset{n}{\Gamma}_{abc} \sum_{a,b,c=1}^{n-1} \bar{q}_{ia}\bar{q}_{jb}\bar{q}_{kc},$$

where $i \neq j$, $k \notin \{i,j\}$ and $a \neq b$, $c \notin \{a,b\}$. Thus,

$$G_m = G_n \left(3 \sum_{a,b=1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{kb} + \sum_{a,b,c=1}^{n-1} \bar{q}_{ia}\bar{q}_{jb}\bar{q}_{kc} \right) = G_n \frac{1}{h^2}.$$

Now, it is easy to see that the value of $\overset{m}{\Gamma}_{ijk}$ at any $x = (r_1/n, \ldots, r_m/n)$, with $\sum_{i=1}^m r_i = n$ and all r_i positive integers, is determined by G_n through (3.12). In fact, the Markov embedding defined by

$$q_{ij} = \begin{cases} 1/r_i & \text{if} \quad j \in R_i \\ 0 & \text{otherwise}, \end{cases}$$

where $R_i = \{r_1 + \cdots + r_{i-1} + 1, \ldots, r_1 + \cdots + r_i\}$, $i = 1, \ldots, m$, maps x to $y = (1/n, \ldots, 1/n)$. Finally, every point in S_{m-1} can be approximated arbitrarily well by an x of the form $(r_1/n, \ldots, r_m/n)$ and, since the $\overset{m}{\Gamma}_{ijk}$'s are C^∞ functions, we obtain the result. □

Following the same steps as in the third order case, we can obtain the characterization of any e-invariant array of fourth order.

THEOREM 5.1. *The only e-invariant fourth order arrays have components of the form*

$$\begin{aligned} \overset{m}{T}_{ijkl}(x) = C\Bigg\{ &\left(\frac{1}{x_m} + \frac{\delta_{il}}{x_i}\right)\left(\frac{1}{x_m} + \frac{\delta_{jk}}{x_j}\right) + \left(\frac{1}{x_m} + \frac{\delta_{jl}}{x_j}\right)\left(\frac{1}{x_m} + \frac{\delta_{ik}}{x_k}\right) \\ &+ \left(\frac{1}{x_m} + \frac{\delta_{kl}}{x_k}\right)\left(\frac{1}{x_m} + \frac{\delta_{ij}}{x_i}\right)\Bigg\} \\ &+ D\left(\frac{1}{x_m^3} + \frac{\delta_{ijkl}}{x_i^3}\right), \end{aligned}$$

$i, j, k, l = 1, \ldots, m-1$, where C and D are constants.

PROOF. By definition (3.14), any e-invariant array of fourth order, T_{ijkl}, should verify

$$\overset{m}{T}_{ijkl}(x) = \sum_{a,b,c,d} \overset{n}{T}_{abcd}(y)\bar{q}_{ia}\bar{q}_{jb}\bar{q}_{kc}\bar{q}_{ld}.$$

Then, if we take $n = m > 3$, $x = (1/m, \ldots, 1/m)$ and the Markov embedding that permutes the first $m-1$ coordinates, the previous condition implies that $\overset{m}{T}$, in x, is completely symmetric and that:

$$\overset{m}{T}_{iiii} = F_m, \quad \overset{m}{T}_{iiij} = G_m, \quad \overset{m}{T}_{iijj} = H_m, \quad \overset{m}{T}_{iijk} = I_m, \quad \overset{m}{T}_{ijkl} = J_m,$$

where i, j, k, l are all different. Now, if we take a Markov embedding that interchanges some coordinate x_i and x_m, we obtain the relationships:

$$\overset{m}{T}_{jjjj} = \overset{m}{T}_{jjjj} - 4\,\overset{m}{T}_{ijjj} + 6\,\overset{m}{T}_{iijj} - 4\,\overset{m}{T}_{iiij} + \overset{m}{T}_{iiii},$$
$$\overset{m}{T}_{iiij} = -\,\overset{m}{T}_{iiij} + \overset{m}{T}_{iiii}$$

and

$$\overset{m}{T}_{ijkl} = -\,\overset{m}{T}_{ijkl} + 3\,\overset{m}{T}_{iikl} - 3\,\overset{m}{T}_{iiil} + \overset{m}{T}_{iiii},$$

that is,

$$F_m = 8G_m - 6H_m,$$
$$F_m = 2G_m$$

and

$$I_m = \frac{2J_m + 3G_m - F_m}{3}.$$

So, the value of $\overset{m}{T}$ in the center of the simplex S_{m-1} is determined by two constants.

The value of $\overset{m}{T}$ in the center of S_{m-1} determines the value of $\overset{n}{T}$ in the center of S_{n-1}, for any integer $n > 1$. In fact, if we consider the Markov embedding (5.1), we obtain

$$\overset{m}{T}_{iiii} = \overset{n}{T}_{aaaa} \sum_{a=1}^{n-1} \bar{q}_{ia}^4 + 4\,\overset{n}{T}_{aaab} \sum_{a,b=1}^{n-1} \bar{q}_{ia}^3\bar{q}_{ib} + 3\,\overset{n}{T}_{aabb} \sum_{a,b=1}^{n-1} \bar{q}_{ia}^2\bar{q}_{ib}^2$$
$$+ 6\,\overset{n}{T}_{aabc} \sum_{a,b,c=1}^{n-1} \bar{q}_{ia}^2\bar{q}_{ib}\bar{q}_{ic} + \overset{n}{T}_{abcd} \sum_{a,b,c,d=1}^{n-1} \bar{q}_{ia}\bar{q}_{ib}\bar{q}_{ic}\bar{q}_{id},$$

and

$$\overset{m}{T}_{ijkl} = \overset{n}{T}_{aaaa} \sum_{a=1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{ka}\bar{q}_{la} + 4\,\overset{n}{T}_{aaab} \sum_{a,b=1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{ka}\bar{q}_{lb}$$
$$+ 3\,\overset{n}{T}_{aabb} \sum_{a,b=1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{kb}\bar{q}_{lb} + 6\,\overset{n}{T}_{aabc} \sum_{a,b,c=1}^{n-1} \bar{q}_{ia}\bar{q}_{ja}\bar{q}_{kb}\bar{q}_{lc}$$
$$+ \overset{n}{T}_{abcd} \sum_{a,b,c,d=1}^{n-1} \bar{q}_{ia}\bar{q}_{jb}\bar{q}_{kc}\bar{q}_{ld},$$

where different indices never take the same value. Thus,

$$\begin{aligned}2G_m &= 2G_n\frac{2h-1}{h^4} - G_n\frac{8(h-1)-6(h-1)(2h-1)}{h^4} \\ &\quad - (G_n+2J_n)\frac{4(h-1)(2h-3)}{h^4} + 12J_n\frac{(h-1)(h-2)}{h^4} \\ &= G_n\frac{4h-2}{h^3} - J_n\frac{4h-4}{h^3}\end{aligned}$$

and

$$\begin{aligned}J_m &= 2G_n\frac{h-1}{h^4} - G_n\frac{8(h-1)}{h^4} + 3G_n\frac{(h-1)(h-2)}{h^4} \\ &\quad - 2(G_n+2J_n)\frac{(h-6)(h-1)}{h^4} + J_n\frac{3h^2-26h+24}{h^4} \\ &= G_n\frac{h-1}{h^3} - J_n\frac{h-2}{h^3}\end{aligned}$$

and we can obtain G_n and J_n from G_m and J_m.

Finally, as in the third order case, the value in the center of the simplex determines the value everywhere. So, the linear space of e-invariant fourth order arrays has dimension two and since

$$\left(\frac{1}{x_m}+\frac{\delta_{il}}{x_i}\right)\left(\frac{1}{x_m}+\frac{\delta_{jk}}{x_j}\right)+\left(\frac{1}{x_m}+\frac{\delta_{jl}}{x_j}\right)\left(\frac{1}{x_m}+\frac{\delta_{ik}}{x_k}\right) + \left(\frac{1}{x_m}+\frac{\delta_{kl}}{x_k}\right)\left(\frac{1}{x_m}+\frac{\delta_{ij}}{x_i}\right)$$

and

$$\left(\frac{1}{x_m^3}+\frac{\delta_{ijkl}}{x_i^3}\right)$$

are e-invariant, we obtain the result. □

References

Amari, S. (1985). *Differential-geometrical Methods in Statistics*, Lecture Notes in Statistics, Vol. 28, Springer, New York.

Blæsild, P. (1991). Yokes and tensors derived from yokes, *Ann. Inst. Statist. Math.*, **43** (1), 95–113.

Burbea, J. (1983). *J*-divergences and related concepts, *Encyclopedia of Statistical Sciences*, Vol. 4, 290–296, Wiley, New York.

Campbell, L. (1986). An extended Čencov characterization of the information metric, *Proc. Amer. Math. Soc.*, **98**, 135–141.

Čencov, N. (1972). *Statistical Decision Rules and Optimal Inference*, Nauka, Moscow (in Russian, English version (1982). *Trans. Math. Monographs*, **53**, American Mathematical Society, Providence).

Eguchi, S. (1992). Geometry of minimum contrast, *Hiroshima Math. J.*, **22**, 631–647.

Heyer, H. (1982). *Theory of Statistical Experiments*, Springer, New York.