CAIXA 15.1

UNIVERSITAT DE BARCELONA

ON THE RELATIONSHIP BETWEEN α-CONNECTIONS AND THE ASYMPTOTIC PROPERTIES OF PREDICTIVE DISTRIBUTIONS

by

J.M. Corcuera and *F. Giummolè*

AMS 1980 Subjects Classifications: 62F10, 62B10, 62A99

Mathematics Preprint Series No. 210
July 1996

On the relationship between α-connections and the asymptotic properties of predictive distributions

J.M. Corcuera *
Departament d'Estadística
Universitat de Barcelona
Gran Via de les Corts Catalanes, 585
08007 Barcelona, Spain

F. Giummolè
Dipartimento di Scienze Statistiche
Università degli Studi di Padova
Via S. Francesco, 33
35121 Padova, Italy

Abstract

In a recent paper Komaki studies the second-order asymptotic properties of the predictive distributions, using the Kullback-Leibler divergence as loss function. He shows that estimative distributions with asymptotically efficient estimators can be improved by predictive distributions that do not belong to the model. The model is assumed to be a multidimensional curved exponential family. In this paper we generalize the result assuming as loss function any f-divergence. It appears a relationship between the α-connections and the optimal predictive distributions. In particular, using an α-divergence to measure the goodness of a predictive distribution, the optimal shift of the estimative distribution is related with α-covariant derivatives. The expression we obtain for the asymptotic risk is also useful to study the higher-order asymptotic properties of an estimator, in the mentioned class of loss functions.

AMS 1980 subjects classifications: 62F10, 62B10, 62A99.

Keywords and phrases: curved exponential family, differential geometry, f-divergences, predictive distributions, second-order asymptotic theory, α-connections, α-embedding curvature.

1 Introduction

The main goal of this work is to provide distributions that are close, in the sense of an f-divergence, to an unknown distribution belonging to a curved exponential family

$$\mathcal{P} = \left\{p(x;\theta(u)) = \exp[\theta^i(u)x_i - \psi(\theta(u))]\right\}.$$

In order to obtain this, we could estimate u by $\hat{u}$ and consider $p(x;\hat{u})$. This kind of distributions are called *estimative distributions.* The procedure assures that they belong to the model. However, perhaps we could obtain a better result by considering *predictive distributions*, that is, distributions outside the model.

Let $\hat{p}(x;x_{1\sim N})$ be a predictive distribution obtained by some rule from the sample of size N, $x_{1\sim N} = (x(1),\ldots,x(N))$. An f-divergence D_f of the predictive distribution to the true one is defined as:

$$D_f\left(p(x;u),\hat{p}(x;x_{1\sim N})\right) = \int f\left(\frac{\hat{p}(x;x_{1\sim N})}{p(x;u)}\right) p(x;u)\mu(dx),$$

where f is a convex function with minimum value in 1. We measure the closeness, by

$$E_u\left(D_f(p,\hat{p})\right) = \int D_f\left(p(x;u),\hat{p}(x;x_{1\sim N})\right) p(x_{1\sim N};u)dx_{1\sim N}. \tag{1}$$

In order to choose $\hat{p}$, we could try to find the distribution that minimizes (1), uniformly in u, among "all probability distributions" equivalent to p. Since there are some technical problems in giving a structure of differentiable manifold to this infinite dimensional space, we follow the procedure suggested by Komaki (1995) and try to solve the problem only for distributions belonging to a finite dimensional model containing $\mathcal{P}$. We construct this model by enlarging $\mathcal{P}$ in orthogonal directions. As we shall see, only a finite number of special directions contribute to improve the estimative distribution, so that the solution does not depend on the enlarged model, whenever it contains such directions; that is, we can add more and more orthogonal directions without changing the solution. In this sense, we can consider the problem solved in the infinite dimensional space $\mathcal{F}$ of all probability distributions equivalent to p. The relevant directions are just the difference between what we call, with certain abuse of language, the α-covariant derivatives of the α-score function in $\mathcal{F}$ and in $\mathcal{P}$.

For the sake of simplicity, we shall work with α-divergences D_α, that is f-divergences with

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

Note that D_α is a continuous function with respect to α. In the final remark, we extend the results to any f-divergence.

2 The enlarged model

Let $\mathcal{E}$ be a n-dimensional full exponential family, that is,

$$\mathcal{E} = \left\{p(x;\theta) = \exp[\theta^i x_i - \psi(\theta)],\, \theta \in \Theta\right\},$$

where the probability functions $p(x;\theta)$ are densities with respect to some reference measure μ and

$$\Theta = \left\{\theta : \int \exp[\theta^i x_i]\mu(dx) < \infty\right\}$$

is an open subset of $\mathbb{R}^n$. We consider the model $\mathcal{P}$ to be a (n,m)-curved exponential family of $\mathcal{E}$, $m \leq n$,

$$\mathcal{P} = \left\{p(x;u) = \exp[\theta^i(u)x_i - \psi(\theta(u))],\, u \in U\right\},$$

with U smooth m-dimensional submanifold of Θ.

Let

$$l_\alpha(x;u) = \begin{cases} \frac{2}{1-\alpha}\left[p^{\frac{1-\alpha}{2}}(x;u) - 1\right] & \alpha \neq 1 \\ \log p(x;u) & \alpha = 1 \end{cases}$$

be the so-called α-representation of $p(x;u)$, see Amari (1985), p.66. From now on, the index α will be used to denote all that regards α-representation of geometric quantities. The tangent space T_u of $\mathcal{P}$ in u is identified with the vector space spanned by

$$\partial_a l_\alpha(x;u) = \frac{\partial l_\alpha(x;u)}{\partial u^a}, \quad a = 1, \cdots, m,$$

that are the components of what we call the α-score function. The first and second derivatives of $l_\alpha(x;u)$ are related to those of $l(x;u) = \log p(x;u) = l_1(x;u)$ by

$$\partial_i l_\alpha = p^{\frac{1-\alpha}{2}} \partial_i l$$

and

$$\partial_i \partial_j l_\alpha = p^{\frac{1-\alpha}{2}} \left(\partial_i \partial_j l + \frac{1-\alpha}{2} \partial_i l \partial_j l\right).$$

Defining

$$E_\alpha(f(x)) = \int f(x) p^\alpha(x;u) \mu(dx),$$

we have that the inner product of vectors $\partial_a l_\alpha$ and $\partial_b l_\alpha$,

$$\langle \partial_a l_\alpha, \partial_b l_\alpha \rangle_\alpha = E_\alpha(\partial_a l_\alpha \partial_b l_\alpha) = \int \partial_a l_\alpha \partial_b l_\alpha p^\alpha \mu(dx) = \int \partial_a l \partial_b l \, p \, \mu(dx) = \langle \partial_a l, \partial_b l \rangle,$$

does not depend on the α-representation; it is the (a,b)-component of the Fisher matrix, g_{ab}. In the sequel, we omit the subscript α in the inner product and in the expectation, since it will be clear from the representation used. We indicate with g^{ab} the inverse of g_{ab} and use the repeated index convention.

Following Amari (Amari et al. 1987), we can construct a fibre bundle on $\mathcal{P}$ by associating to each point $p(x;u) \in \mathcal{P}$ a linear space H_u defined by

$$H_u = \left\{h(x) : \int p^{\frac{1+\alpha}{2}}(x;u)\, h(x) \mu(dx) = 0, \int p^\alpha(x;u)\, h^2(x) \mu(dx) < \infty\right\}$$

If $h,\ g \in H_u$ we can define an inner product on H_u by

$$\langle h, g\rangle = \int p^{\alpha}(x;u)h(x)g(x)\mu(dx);$$

it is well defined by the Cauchy-Schwarz inequality. Then, since H_u is a closed subset of $L^2(p^{\alpha}\mu)$, it is a Hilbert space. The tangent vectors $\partial_a l_\alpha(x;u)$ satisfy

$$\int p^{\frac{1+\alpha}{2}}(x;u)\,\partial_a l_\alpha(x;u)\mu(dx) = \int p(x;u)\,\partial_a l(x;u)\mu(dx) = 0$$

and

$$\int p^{\alpha}(x;u)\,(\partial_a l_\alpha(x;u))^2\mu(dx) = \int p(x;u)\,(\partial_a l(x;u))^2\mu(dx) = g_{aa} < \infty;$$

thus, $T_u \subset H_u$. Notice that the inner product defined on T_u is compatible with that in H_u. Attached to each point we have a different Hilbert space and the aggregate

$$\mathcal{H}(\mathcal{P}) = \bigcup_{u\in U} H_u$$

constitutes the fibre bundle. It is necessary to establish a one to one correspondence between H_u and $H_{u'}$, when $p(x;u)$ and $p(x;u')$ are neighbouring points, in order to express the rate of variation of a vector field as an element of the fibre bundle. If we move in the direction $\partial_a l_\alpha$ and $h_u \in H_u$, $\partial_a h_u \notin H_u$ in general. Anyway, if h_u is a smooth vector field, in the sense that we can interchange the integral and the derivative,

$$\begin{aligned}
0 &= \partial_a \int p^{\frac{1+\alpha}{2}}\, h_u\mu(dx) \\
&= \int \frac{1+\alpha}{2} p^{\frac{\alpha-1}{2}}\,\partial_a p\, h_u\mu(dx) + \int p^{\frac{1+\alpha}{2}}\,\partial_a h_u\mu(dx) \\
&= \frac{1+\alpha}{2}\int p^{\frac{\alpha+1}{2}}\,\partial_a l\, h_u\mu(dx) + \int p^{\frac{1+\alpha}{2}}\,\partial_a h_u\mu(dx) \\
&= \frac{1+\alpha}{2}\int p^{\alpha}\,\partial_a l_\alpha\, h_u\mu(dx) + \int p^{\frac{1+\alpha}{2}}\,\partial_a h_u\mu(dx) \\
&= \int p^{\frac{1+\alpha}{2}}\left[\partial_a h_u + \frac{1+\alpha}{2} p^{\frac{1-\alpha}{2}} E(\partial_a l_\alpha\, h_u)\right]\mu(dx).
\end{aligned}$$

Thus, we can define the α-covariant derivative in $\mathcal{H}$ as:

$$\overset{\alpha}{\nabla}{}^{(\mathcal{H})}_{\partial_a l_\alpha} h_u = \partial_a h_u + \frac{1+\alpha}{2} p^{\frac{1-\alpha}{2}} E(\partial_a l_\alpha h_u).$$

If $h_u(x) = \partial_b l_\alpha(x;u)$ (notice that it is a smooth vector field), we have

$$\overset{\alpha}{\nabla}{}^{(\mathcal{H})}_{\partial_a l_\alpha} \partial_b l_\alpha = \partial_a \partial_b l_\alpha + \frac{1+\alpha}{2} p^{\frac{1-\alpha}{2}} g_{ab}$$

and the α-covariant derivative in $\mathcal{P}$ is the projection of $\overset{\alpha}{\nabla}{}^{(\mathcal{H})}_{\partial_a l_\alpha} \partial_b l_\alpha$ on T_u:

$$\overset{\alpha}{\nabla}_{\partial_a l_\alpha} \partial_b l_\alpha = \langle \overset{\alpha}{\nabla}{}^{(\mathcal{H})}_{\partial_a l_\alpha} \partial_b l_\alpha, \partial_c l_\alpha \rangle g^{cd} \partial_d l_\alpha = \overset{\alpha}{\Gamma}_{abc}\, g^{cd} \partial_d l_\alpha.$$

These connections coincides with the α-connections defined in Amari (1985), p.38. It is natural to define

$$\overset{\alpha}{\nabla}{}^{(\mathcal{F})}_{\partial_a l_\alpha} \partial_b l_\alpha = \overset{\alpha}{\nabla}{}^{(\mathcal{H})}_{\partial_a l_\alpha} \partial_b l_\alpha,$$

even though we do not have an α-covariant derivative in the whole $\mathcal{F}$. We use the superscripts m and e respectively for the -1 and $+1$-covariant derivatives.

Let $\mathcal{M}$ be any regular parametric model containing $\mathcal{P}$. We can consider on $\mathcal{M}$ the coordinate system (u,s), where u^a, $a = 1,\ldots,m$, is the old coordinate system on $\mathcal{P}$ and s^I, $I = m+1,\ldots,r$, $r > m$, are orthogonal coordinates to $\mathcal{P}$. Moreover we suppose $s = 0$ for the points in $\mathcal{P}$. The tangent space to the enlarged model $\mathcal{M}$ is now spanned by vectors $\partial_a l_\alpha(x;u,s)$, $a = 1,\ldots,m$, and $\partial_I l_\alpha(x;u,s)$, $I = m+1,\ldots,r$. Omitting the argument $(x;u,s)$ will not cause any confusion since we are interested on the tangent space to $\mathcal{M}$ on the points with coordinates $s = 0$ and

$$\partial_a l_\alpha(x;u,s)|_{s=0} = \partial_a l_\alpha(x;u), \quad a = 1,\ldots,m.$$

We call h_I the tangent vectors $\partial_I l_\alpha(x;u,s)|_{s=0}$, $I = m+1,\ldots,r$. Notice that the h_I's belong to H_u. If $s^I = O(N^{-1})$, we can write

$$p(x;u,s) = p(x;u) + p^{\frac{1+\alpha}{2}}(x;u) s^I h_I(x) + o(N^{-1}), \tag{2}$$

since

$$h_I(x) = \partial_I l_\alpha(x; u, s)|_{s=0} = p^{-\frac{1+\alpha}{2}}(x; u)\partial_I p(x; u, s)|_{s=0}.$$

Expression (2) is an approximation, up to order N^{-1}, for the predictive distribution. It integrates one, since vectors h_I's belong to H_u. In the case when

$$\int p(x; u) \exp\left[\frac{s^I h_I(x)}{p^{\frac{1-\alpha}{2}}(x; u)}\right] \mu(dx) < \infty,$$

we can obtain another useful expression for $p(x; u, s)$. Since the h_I's belong to H_u, we have

$$\int p(x; u) \exp\left[\frac{s^I h_I(x)}{p^{\frac{1-\alpha}{2}}(x; u)}\right] \mu(dx) = 1 + o(N^{-1})$$

and

$$p(x; u, s) = p(x; u) \exp\left[\frac{s^I h_I(x)}{p^{\frac{1-\alpha}{2}}(x; u)} - \phi_u(s)\right] + o(N^{-1}), \tag{3}$$

where

$$\phi_u(s) = \log \int p(x; u) \exp\left[\frac{s^I h_I(x)}{p^{\frac{1-\alpha}{2}}(x; u)}\right] \mu(dx).$$

3 Predictive distribution

We consider predictive distributions $p(x; \hat{u}_N(\bar{x}), \hat{s}(\bar{x}))$, with $\hat{u}_N(\bar{x})$ a smooth asymptotically efficient estimator, hence first order equivalent to the maximum likelihood estimator, of the form

$$\hat{u}_N(\bar{x}) = \hat{u}_\infty(\bar{x}) + \frac{1}{N}\bar{u}(\bar{x}) + o_p(N^{-1}), \tag{4}$$

where, fixed $\bar{x}$,

$$\hat{u}_\infty(\bar{x}) = \lim_{N\to\infty} \hat{u}_N(\bar{x})$$

and

$$\bar{u}(\bar{x}) = \lim_{N\to\infty} N\left(\hat{u}_N(\bar{x}) - \hat{u}_\infty(\bar{x})\right)$$

depend on N only through $\bar{x}$.

For each N, $\hat{u}_N$ is a map

$$\hat{u}_N : \mathcal{E} \to \mathcal{P},$$

since $\bar{x}$ can be identified with the point in $\mathcal{E}$ having expectation parameters $\eta_i = \bar{x}_i$. Then, $\hat{u}_\infty$ is also a map from $\mathcal{E}$ to $\mathcal{P}$ and we can associate to $\hat{u}_N$ a family of ancillary $(n-m)$-dimensional submanifolds of $\mathcal{E}$, $\mathcal{A} = \{A(u)\}$, where $A(u) = \hat{u}_\infty^{-1}(u)$. Following Amari (1985), p.128, it can be shown that $\hat{u}_\infty$ is consistent if and only if every $p(x;u) \in \mathcal{P}$ is contained in the associated submanifold $A(u)$ and $\hat{u}_\infty$ is asymptotically first order efficient if and only if $A(u)$ is orthogonal to $\mathcal{P}$ in u. On the other hand, since

$$\lim_{N\to\infty} \hat{u}_\infty(\bar{x}) = \lim_{N\to\infty} \hat{u}_N(\bar{x})$$

in probability and

$$\lim_{N\to\infty} \left[\sqrt{N}(\hat{u}_\infty(\bar{x}) - u)\right] = \lim_{N\to\infty} \left[\sqrt{N}(\hat{u}_N(\bar{x}) - u)\right]$$

in distribution, the results still hold for $\hat{u}_N$.

If we introduce a coordinate system v^κ, $\kappa = m+1, \ldots, n$ on each $A(u)$, every point in the full exponential family containing $\mathcal{P}$ is uniquely determined by a pair (u, v). It is convenient to fix $v = 0$ for the points in $\mathcal{P}$. We denote by indices $a, b, c, \ldots \in \{1, \ldots, m\}$ the coordinates u in $\mathcal{P}$, by $\kappa, \lambda, \mu, \ldots \in \{m+1, \ldots, n\}$ the coordinates v in $A(u)$ and by $\alpha, \beta, \gamma, \ldots \in \{1, \ldots, n\}$ the new coordinates $w = (u, v)$ in $\mathcal{E}$. Since $\hat{u}_N$ is asymptotically efficient,

$$g_{a\kappa}(u) = 0.$$

Indices $i, j, \ldots \in \{1, \ldots, n\}$ are used to denote both the natural parameters θ and the expectation parameters η in $\mathcal{E}$. We use indices $I, J, K, \ldots \in \{m+1, \ldots, r\}$ for the coordinates s we add to enlarge the model $\mathcal{P}$ and $A, B, C, \ldots \in \{1, \ldots, r\}$ for the coordinates $t = (u, s)$ in the enlarged model $\mathcal{M}$. By the coordinate system we choose on $\mathcal{M}$,

$$g_{aI}(u) = 0.$$

Theorem 3.1 *The average α-divergence from the true distribution $p(x;u_0)$ to a predictive distribution $p(x;\hat{u}_N(\bar{x}),\hat{s}(\bar{x}))$ is given by*

$$
\begin{aligned}
&E_{u_0}\left\{D_\alpha\left(p(x;u_0),p(x;\hat{u}_N(\bar{x}),\hat{s}(\bar{x}))\right)\right\} = \qquad (5)\\
&= \frac{m}{2N}+\frac{1}{4N^2}\left[2\left(\overset{e}{H}{}^2_{\mathcal{P}}\right)+\left(\overset{m}{H}{}^2_{\mathcal{A}}\right)\right]\\
&+\frac{1}{2N^2}g_{ab}\left(\bar{u}^a-\frac{1}{2}\overset{m}{H}_{\kappa\lambda}{}^a g^{\kappa\lambda}\right)\left(\bar{u}^b-\frac{1}{2}\overset{m}{H}_{\mu\nu}{}^b g^{\mu\nu}\right)\\
&+\frac{1}{N^2}\overset{\frac{1-\alpha}{2}}{\nabla}_a\left(\bar{u}^a-\frac{1}{2}\overset{m}{H}_{\kappa\lambda}{}^a g^{\kappa\lambda}\right)\\
&+\frac{1}{2N^2}\left(g_{IJ}\bar{s}^I\bar{s}^J-\overset{\alpha}{H}_{abI}\,g^{ab}\bar{s}^I\right)\\
&+\frac{\alpha-3}{12N^2}T_{abc}T^{abc}+\frac{(\alpha-11)(\alpha-1)}{32N^2}Q_{abcd}g^{ab}g^{cd}\\
&+\frac{1}{4N^2}g^{ac}g^{bd}\int\left(\partial_a\partial_b p-\overset{m}{\Gamma}_{ab}{}^e\partial_e p\right)\left(\partial_c\partial_d p-\overset{m}{\Gamma}_{cd}{}^f\partial_f p\right)\frac{1}{p}\mu(dx)\\
&-\frac{3}{8N^2}g^{ab}g^{cd}\int\left(\partial_a\partial_b p-\overset{m}{\Gamma}_{ab}{}^e\partial_e p\right)\left(\partial_c\partial_d p-\overset{m}{\Gamma}_{cd}{}^f\partial_f p\right)\frac{1}{p}\mu(dx)\\
&-\frac{1}{N^2}g^{ac}g^{bd}\int\partial_a p\,\partial_b p\left(\partial_c\partial_d p-\overset{m}{\Gamma}_{cd}{}^f\partial_f p\right)\frac{1}{p^2}\mu(dx)\\
&+\frac{\alpha+1}{8N^2}g^{ab}g^{cd}\overset{m}{\nabla}_d T_{abc}+o(N^{-2}),
\end{aligned}
$$

where all the quantities are evaluated in u_0,

$$
\begin{aligned}
\bar{s} &= NE\left[\hat{s}(\bar{x})\right],\\
Q_{abcd} &= E\left(\partial_a l\partial_b l\partial_c l\partial_d l\right),\\
\overset{\alpha}{H}_{rst} &= \left\langle \overset{\alpha}{\nabla}_{\partial_r l_\alpha}\,\partial_s l_\alpha,\partial_t l_\alpha\right\rangle,\\
T_{abc} &= E(\partial_a l\partial_b l\partial_c l),\\
\left(\overset{e}{H}{}^2_{\mathcal{P}}\right) &= \overset{e}{H}{}^{ac\kappa}\,\overset{e}{H}{}^{bd\lambda}g_{cd}g_{\kappa\lambda}g_{ab},\\
\left(\overset{m}{H}{}^2_{\mathcal{A}}\right) &= \overset{m}{H}{}^{\kappa\lambda a}\,\overset{m}{H}{}^{\mu\nu b}g_{\kappa\mu}g_{\lambda\nu}g_{ab}
\end{aligned}
$$

and $\overset{\alpha}{\nabla}_a$ is the a-component of the general covariant derivative of a tensor with respect to the α-connection.

Proof: For simplicity, we omit the subscript N and write $\hat{u}(\bar{x})$ for $\hat{u}_N(\bar{x})$. By expanding an α-divergence from $p(x;u_0)$ to $p(x;\hat{u},\hat{s})$, we obtain:

$$
\begin{aligned}
D_\alpha(p(x;u_0),p(x;\hat{u},\hat{s})) &= \int f_\alpha\left(\frac{p(x;\hat{u},\hat{s})}{p(x;u_0)}\right) p(x;u_0)\mu(dx) \\
&= f_\alpha(1) + E_{u_0}(\partial_A f_\alpha)\tilde{t}^A + \frac{1}{2}E_{u_0}(\partial_A\partial_B f_\alpha)\tilde{t}^A\tilde{t}^B \\
&\quad + \frac{1}{6}E_{u_0}(\partial_A\partial_B\partial_C f_\alpha)\tilde{t}^A\tilde{t}^B\tilde{t}^C + \frac{1}{24}E_{u_0}(\partial_A\partial_B\partial_C\partial_D f_\alpha)\tilde{t}^A\tilde{t}^B\tilde{t}^C\tilde{t}^D + o(|\tilde{t}|^4) \\
&= \frac{1}{2}g_{AB}(u_0)\tilde{t}^A\tilde{t}^B + \left(\frac{1}{2}\overset{\alpha}{\Gamma}_{ABC}(u_0) + \frac{\alpha}{3}T_{ABC}(u_0)\right)\tilde{t}^A\tilde{t}^B\tilde{t}^C \\
&\quad + K_{ABCD}(u_0)\tilde{t}^A\tilde{t}^B\tilde{t}^C\tilde{t}^D + o(|\tilde{t}|^4),
\end{aligned}
$$

where $\tilde{t} = \hat{t} - t_0 = (\hat{u} - u_0, \hat{s})$ and $K_{ABCD} = \frac{1}{24}E_{u_0}(\partial_A\partial_B\partial_C\partial_D f_\alpha)$. Taking into account that, from the definition of f_α,

$$
f_\alpha(1) = 0, \quad f''_\alpha(1) = 1, \quad f'''_\alpha(1) = \frac{\alpha-3}{2}, \quad f^{(4)}_\alpha(1) = \frac{(\alpha-3)(\alpha-5)}{4},
$$

we can write K_{ABCD} in a form that will be useful for the calculations:

$$
\begin{aligned}
K_{ABCD} = \frac{1}{24}\Bigg\{&\frac{(\alpha-3)(\alpha-5)}{4}\int\frac{\partial_A p\partial_B p\partial_C p\partial_D p}{p^3}\mu(dx) \\
&+ \frac{\alpha-3}{2}\int\frac{\partial_A\partial_B p\partial_C p\partial_D p}{p^2}\mu(dx)[6] + \int\frac{\partial_A\partial_B\partial_C p\partial_D p}{p}\mu(dx)[4] \\
&+ \int\frac{\partial_A\partial_B p\partial_C\partial_D p}{p}\mu(dx)[3]\Bigg\},
\end{aligned} \tag{6}
$$

where the bracket [ ] refers to the sum of a number of different terms obtained by permutation of the indices. We suppose $\hat{s}(\bar{x})$ to be a smooth function of $\bar{x}$ and $O_p(N^{-1})$. Since $g_{aI}(u_0) = 0$, $\overset{\alpha}{\Gamma}_{ABC}$ is symmetric with respect to indices A and B and T_{ABC} is a symmetric tensor, we can rewrite the expansion of D_α as:

$$
\begin{aligned}
&D_\alpha(p(x;u_0),p(x;\hat{u},\hat{s})) = \\
&= \frac{1}{2}g_{ab}(u_0)\tilde{u}^a\tilde{u}^b + \frac{1}{2}g_{IJ}(u_0)\hat{s}^I\hat{s}^J + \left(\frac{1}{2}\overset{\alpha}{\Gamma}_{abc}(u_0) + \frac{\alpha}{3}T_{abc}(u_0)\right)\tilde{u}^a\tilde{u}^b\tilde{u}^c \\
&\quad + \left[\frac{1}{2}\left(\overset{\alpha}{\Gamma}_{abI}(u_0) + 2\,\overset{\alpha}{\Gamma}_{aIb}(u_0)\right) + \alpha T_{abI}(u_0)\right]\tilde{u}^a\tilde{u}^b\hat{s}^I + K_{abcd}(u_0)\tilde{u}^a\tilde{u}^b\tilde{u}^c\tilde{u}^d \\
&\quad + o_p(N^{-2}),
\end{aligned}
$$

where $\tilde{u} = \hat{u} - u_0$. The mean value of D_α is:

$$
\begin{aligned}
&E_{u_0}\{D_\alpha(p(x;u_0),p(x;\hat{u},\hat{s}))\} = \qquad (7)\\
&= \frac{1}{2}g_{ab}(u_0)E_{u_0}[\tilde{u}^a\tilde{u}^b]\\
&+\frac{1}{2}g_{IJ}(u_0)E_{u_0}[\hat{s}^I\hat{s}^J]\\
&+\left(\frac{1}{2}\overset{\alpha}{\Gamma}_{abc}(u_0)+\frac{\alpha}{3}T_{abc}(u_0)\right)E_{u_0}[\tilde{u}^a\tilde{u}^b\tilde{u}^c]\\
&+\left(\frac{1}{2}\overset{\alpha}{\Gamma}_{abI}(u_0)+\overset{\alpha}{\Gamma}_{aIb}(u_0)+\alpha T_{abI}(u_0)\right)E_{u_0}[\tilde{u}^a\tilde{u}^b\hat{s}^I]\\
&+K_{abcd}(u_0)E_{u_0}[\tilde{u}^a\tilde{u}^b\tilde{u}^c\tilde{u}^d]+o(N^{-2}).
\end{aligned}
$$

For the calculations we use the following relations:

$$\overset{\alpha}{\Gamma}_{abc}=\overset{m}{\Gamma}_{abc}-\frac{1+\alpha}{2}T_{abc},$$

$$\overset{-\alpha}{\Gamma}_{abc}=\overset{\alpha}{\Gamma}_{abc}+\alpha T_{abc}.$$

Moreover, since $g_{b\kappa}=0$, we have that

$$
\begin{aligned}
0=\partial_a g_{b\kappa} &= \int\frac{\partial_a\partial_b p\partial_\kappa p}{p}\mu(dx)+\int\frac{\partial_b p\partial_a\partial_\kappa p}{p}\mu(dx)-\int\frac{\partial_a p\partial_b p\partial_\kappa p}{p^2}\mu(dx)\\
&= \overset{m}{\Gamma}_{ab\kappa}+\overset{m}{\Gamma}_{a\kappa b}-T_{ab\kappa}\\
&= \overset{\alpha}{\Gamma}_{ab\kappa}+\overset{\alpha}{\Gamma}_{a\kappa b}+\alpha T_{ab\kappa}.
\end{aligned}
$$

It follows that

$$\overset{\alpha}{\Gamma}_{a\kappa b}=-\overset{\alpha}{\Gamma}_{ab\kappa}-\alpha T_{ab\kappa}$$

and, similarly,

$$\overset{\alpha}{\Gamma}_{aIb}=-\overset{\alpha}{\Gamma}_{abI}-\alpha T_{abI}.$$

Let us begin by calculating $E_{u_0}[\tilde{u}^a\tilde{u}^b]$. First of all notice that, defining

$$\bar{l}=\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\log p(x(l);u)=\sqrt{N}\{\theta^i(u)\bar{x}_i-\psi(\theta(u))\},$$

we obtain

$$\tilde{x}_i = \bar{x}_i - \eta_i = \bar{x}_i - \partial_i \psi = \frac{1}{\sqrt{N}} \partial_i \bar{l}$$

and we can easily calculate the moments of $\tilde{x}$:

$$\begin{aligned}
&E_{u_0}[\tilde{x}_i] = 0, \qquad E_{u_0}[\tilde{x}_i \tilde{x}_j] = \frac{E_{u_0}[\partial_i \bar{l} \partial_j \bar{l}]}{N} = \frac{g_{ij}}{N}, \\
&E_{u_0}[\tilde{x}_i \tilde{x}_j \tilde{x}_k] = \frac{E_{u_0}[\partial_i \bar{l} \partial_j \bar{l} \partial_k \bar{l}]}{N^{\frac{3}{2}}} = \frac{1}{N^2} T_{ijk}, \\
&E_{u_0}[\tilde{x}_i \tilde{x}_j \tilde{x}_k \tilde{x}_h] = \frac{E_{u_0}[\partial_i \bar{l} \partial_j \bar{l} \partial_k \bar{l} \partial_h \bar{l}]}{N^2} = \frac{3}{N^2} g_{(ij} g_{kh)} + O(N^{-3}).
\end{aligned} \tag{8}$$

The bracket ( ) means symmetrization with respect to the indices included, e.g.,

$$3 g_{(ij} g_{kh)} = g_{ij} g_{kh} + g_{ik} g_{jh} + g_{ih} g_{jk}.$$

Since

$$\begin{aligned}
&E_{u_0}[(\tilde{u}^a - g^{ac} \tilde{x}_c)(\tilde{u}^b - g^{bd} \tilde{x}_d)] = \\
&= E_{u_0} \left[\left(\tilde{u}^a - \frac{1}{\sqrt{N}} g^{ac} \partial_c \bar{l} \right) \left(\tilde{u}^b - \frac{1}{\sqrt{N}} g^{bd} \partial_d \bar{l} \right) \right] \\
&= E_{u_0}[\tilde{u}^a \tilde{u}^b] + \frac{1}{N} g^{ab} - \frac{2}{\sqrt{N}} E_{u_0} \left[\partial_d \bar{l} \tilde{u}^{(a} \right] g^{b)d} \\
&= E_{u_0}[\tilde{u}^a \tilde{u}^b] + \frac{1}{N} g^{ab} - \frac{2}{\sqrt{N}} E_{u_0} \left[\partial_d \bar{l} \hat{u}^{(a} \right] g^{b)d} \\
&= E_{u_0}[\tilde{u}^a \tilde{u}^b] + \frac{1}{N} g^{ab} - \frac{2}{N} \partial_d E_{u_0} \left[\hat{u}^{(a} \right] g^{b)d} \\
&= E_{u_0}[\tilde{u}^a \tilde{u}^b] + \frac{1}{N} g^{ab} - \frac{2}{N} \partial_d E_{u_0} \left[\tilde{u}^{(a} + u_0^{(a} \right] g^{b)d} \\
&= E_{u_0}[\tilde{u}^a \tilde{u}^b] - \frac{1}{N} g^{ab} - \frac{2}{N} \partial_c \hat{u}_{bias}^{(a} g^{b)c},
\end{aligned}$$

we can write the mean squared error of $\hat{u}$ as:

$$E_{u_0}[\tilde{u}^a \tilde{u}^b] = \frac{1}{N} g^{ab} + \frac{2}{N} \partial_c \hat{u}_{bias}^{(a} g^{b)c} + E_{u_0}[(\tilde{u}^a - g^{ac} \tilde{x}_c)(\tilde{u}^b - g^{bd} \tilde{x}_d)]. \tag{9}$$

Let $\tilde{w}_\infty = (\tilde{u}_\infty, \hat{v}) = (\hat{u}_\infty - u_0, \hat{v})$. $\bar{x}_i = \eta_i(\hat{u}, \hat{v})$, can be expanded at $(u_0, 0)$ as

$$\bar{x}_i = \eta_i(u_0) + \partial_\alpha \eta_i(u_0)\tilde{w}^\alpha_\infty + \frac{1}{2}\partial_\alpha\partial_\beta\eta_i(u_0)\tilde{w}^\alpha_\infty\tilde{w}^\beta_\infty + O_p(N^{-\frac{3}{2}}).$$

If we put $B_{\alpha i} = \partial_\alpha\eta_i(u_0)$ and $C_{\alpha\beta i} = \partial_\alpha\partial_\beta\eta_i(u_0)$, we can write:

$$\tilde{x}_i = B_{\alpha i}\tilde{w}^\alpha_\infty + \frac{1}{2}C_{\alpha\beta i}\tilde{w}^\alpha_\infty\tilde{w}^\beta_\infty + O_p(N^{-\frac{3}{2}})$$

and, by inversion,

$$\begin{aligned}\tilde{w}^\alpha_\infty &= B^{\alpha i}\tilde{x}_i - \frac{1}{2}B^{\alpha k}B^{\gamma i}B^{\delta j}C_{\gamma\delta k}\tilde{x}_i\tilde{x}_j + O_p(N^{-\frac{3}{2}}) \\ &= g^{\alpha\beta}B_\beta{}^i\tilde{x}_i - \frac{1}{2}C^{\gamma\delta\alpha}B_\gamma{}^iB_\delta{}^j\tilde{x}_i\tilde{x}_j + O_p(N^{-\frac{3}{2}}),\end{aligned}$$

where $B^{\alpha i}$ is the inverse of $B_{\alpha i}$, indices are raised and lowered by multiplication by $g^{\alpha i}$ and $g_{\alpha i}$ and $C_{\beta\gamma}{}^\alpha = C_{\beta\gamma k}B^{\alpha k}$. Since $g^{a\kappa} = 0$, we can write

$$\tilde{u}^a_\infty = g^{ab}B_b{}^i\tilde{x}_i - \frac{1}{2}C^{\gamma\delta a}B_\gamma{}^iB_\delta{}^j\tilde{x}_i\tilde{x}_j + O_p(N^{-\frac{3}{2}}).$$

Notice that

$$B_\alpha{}^i\tilde{x}_i = \frac{1}{\sqrt{N}}B_\alpha{}^i\partial_i\bar{l} = \frac{1}{\sqrt{N}}\partial_\alpha\bar{l} = \tilde{x}_\alpha,$$

since

$$B_\alpha{}^i = B_{\alpha j}g^{ij} = \partial_\alpha\eta_j g^{ij} = \partial_\alpha\partial_j\psi g^{ij} = \frac{\partial\theta^k}{\partial w^\alpha}\partial_k\partial_j\psi g^{ij} = \frac{\partial\theta^i}{\partial w^\alpha}.$$

Moreover,

$$C_{\alpha\beta\gamma} = \overset{m}{\Gamma}_{\alpha\beta\gamma}\,.$$

This is easily prooved because, since the coordinate system η has the property of being flat with respect to the -1-connection, we have that $\overset{m}{\nabla}_{\partial^i}\,\partial^j = 0$, where $\partial^i = \dfrac{\partial}{\partial\eta_i}$, hence $\overset{m}{\nabla}_{\partial_\alpha}\,\partial^j = 0$. Thus,

$$\overset{m}{\nabla}_{\partial_\alpha}\,\partial_\beta = \overset{m}{\nabla}_{\partial_\alpha}\,B_{\beta j}\partial^j = (\partial_\alpha B_{\beta j})\partial^j$$

and, by taking the inner product of each member with $\partial_\gamma = B_{\gamma i}\partial^i$, we obtain the result. Thus we can rewrite the expansion for $\tilde{u}^a_\infty$ as:

$$\tilde{u}^a_\infty = g^{ab}\tilde{x}_b - \frac{1}{2}\overset{m}{\Gamma}{}^{\gamma\delta a}\tilde{x}_\gamma\tilde{x}_\delta + O_p(N^{-\frac{3}{2}}).$$

By (4),

$$\tilde{u}^a = g^{ab}\tilde{x}_b - \frac{1}{2}\overset{m}{\Gamma}{}^{\gamma\delta a}\tilde{x}_\gamma\tilde{x}_\delta + \frac{1}{N}\bar{u}^a(\bar{x}) + o_p(N^{-1}). \tag{10}$$

We can now calculate the bias of $\hat{u}$. By (10) and relations (8),

$$\begin{aligned} \hat{u}^a_{bias} &= E_{u_0}[\hat{u}^a(\bar{x})] - u^a_0 = E_{u_0}[\tilde{u}^a(\bar{x})] \qquad (11)\\ &= -\frac{1}{2}\overset{m}{\Gamma}{}^{\gamma\delta a}E_{u_0}[\tilde{x}_\gamma\tilde{x}_\delta] + \frac{1}{N}E_{u_0}[\bar{u}^a(\bar{x})] + o(N^{-1})\\ &= -\frac{1}{2N}\overset{m}{\Gamma}{}_{bc}{}^a g^{bc} - \frac{1}{2N}\overset{m}{H}{}_{\kappa\lambda}{}^a g^{\kappa\lambda} + \frac{1}{N}\bar{u}^a + o(N^{-1}), \end{aligned}$$

where $\bar{u} = \bar{u}(u_0)$ and $\overset{m}{H}{}_{\kappa\lambda a} = \overset{m}{\Gamma}{}_{\kappa\lambda a}$. By substituting (10) and (11) in (9), we obtain:

$$\begin{aligned} &E_{u_0}[\tilde{u}^a\tilde{u}^b] = \\ &= \frac{1}{N}g^{ab} - \frac{1}{N^2}g^{c(b}\partial_c\left(\overset{m}{\Gamma}{}_{de}{}^{a)}g^{de}\right) - \frac{1}{N^2}g^{c(b}\partial_c\left(\overset{m}{H}{}_{\kappa\lambda}{}^{a)}g^{\kappa\lambda}\right) + \frac{2}{N^2}g^{c(b}\partial_c\bar{u}^{a)}\\ &\quad + E_{u_0}\left[\left(-\frac{1}{2}\overset{m}{\Gamma}{}^{\alpha\beta a}\tilde{x}_\alpha\tilde{x}_\beta + \frac{1}{N}\bar{u}^a(\bar{x})\right)\left(-\frac{1}{2}\overset{m}{\Gamma}{}^{\gamma\delta b}\tilde{x}_\gamma\tilde{x}_\delta + \frac{1}{N}\bar{u}^b(\bar{x})\right)\right] + o(N^{-2}). \end{aligned}$$

Since

$$\begin{aligned} &E_{u_0}\left[\left(-\frac{1}{2}\overset{m}{\Gamma}{}^{\alpha\beta a}\tilde{x}_\alpha\tilde{x}_\beta + \frac{1}{N}\bar{u}^a(\bar{x})\right)\left(-\frac{1}{2}\overset{m}{\Gamma}{}^{\gamma\delta b}\tilde{x}_\gamma\tilde{x}_\delta + \frac{1}{N}\bar{u}^b(\bar{x})\right)\right] = \\ &= E_{u_0}\left[\left(-\frac{1}{2}\overset{m}{\Gamma}{}^{\alpha\beta a}\tilde{x}_\alpha\tilde{x}_\beta + \frac{1}{N}\bar{u}^a\right)\left(-\frac{1}{2}\overset{m}{\Gamma}{}^{\gamma\delta b}\tilde{x}_\gamma\tilde{x}_\delta + \frac{1}{N}\bar{u}^b\right)\right] + o(N^{-2})\\ &= \frac{3}{4N^2}\overset{m}{\Gamma}{}^{\alpha\beta a}\overset{m}{\Gamma}{}^{\gamma\delta b}g_{(\alpha\beta}g_{\gamma\delta)} + \frac{1}{N^2}\bar{u}^a\bar{u}^b\\ &\quad -\frac{1}{N^2}\bar{u}^{(a}\overset{m}{\Gamma}{}_{cd}{}^{b)}g^{cd} - \frac{1}{N^2}\bar{u}^{(a}\overset{m}{H}{}_{\kappa\lambda}{}^{b)}g^{\kappa\lambda} + o(N^{-2})\\ &= \frac{1}{4N^2}\overset{m}{\Gamma}{}^{\alpha\beta a}\overset{m}{\Gamma}{}^{\gamma\delta b}(g_{\alpha\beta}g_{\gamma\delta} + 2g_{\alpha\gamma}g_{\beta\delta}) + \frac{1}{N^2}\bar{u}^a\bar{u}^b \end{aligned}$$

$$-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{\Gamma}_{cd}{}^{b)}g^{cd}-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{H}_{\kappa\lambda}{}^{b)}g^{\kappa\lambda}+o(N^{-2})$$
$$=\frac{1}{4N^2}\left(\overset{m}{\Gamma}_{cd}{}^{a}g^{cd}+\overset{m}{H}_{\kappa\lambda}{}^{a}g^{\kappa\lambda}\right)\left(\overset{m}{\Gamma}_{ef}{}^{b}g^{ef}+\overset{m}{H}_{\mu\nu}{}^{b}g^{\mu\nu}\right)+\frac{1}{2N^2}\overset{m}{\Gamma}{}^{\alpha\beta a}\overset{m}{\Gamma}{}^{\gamma\delta b}g_{\alpha\gamma}g_{\beta\delta}$$
$$+\frac{1}{N^2}\bar{u}^a\bar{u}^b-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{\Gamma}_{cd}{}^{b)}g^{cd}-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{H}_{\kappa\lambda}{}^{b)}g^{\kappa\lambda}+o(N^{-2})$$
$$=\frac{1}{4N^2}\left(\overset{m}{\Gamma}_{cd}{}^{a}g^{cd}+\overset{m}{H}_{\kappa\lambda}{}^{a}g^{\kappa\lambda}\right)\left(\overset{m}{\Gamma}_{ef}{}^{b}g^{ef}+\overset{m}{H}_{\mu\nu}{}^{b}g^{\mu\nu}\right)+\frac{1}{2N^2}\overset{m}{\Gamma}_{cd}{}^{a}\overset{m}{\Gamma}_{ef}{}^{b}g^{ce}g^{df}$$
$$+\frac{1}{2N^2}\left(\overset{m}{H}{}^{\kappa\lambda a}\overset{m}{H}{}^{\mu\nu b}g_{\kappa\mu}g_{\lambda\nu}+2\overset{m}{H}{}^{c\kappa a}\overset{m}{H}{}^{d\lambda b}g_{cd}g_{\kappa\lambda}\right)$$
$$+\frac{1}{N^2}\bar{u}^a\bar{u}^b-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{\Gamma}_{cd}{}^{b)}g^{cd}-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{H}_{\kappa\lambda}{}^{b)}g^{\kappa\lambda}+o(N^{-2})$$
$$=\frac{1}{4N^2}\left(\overset{m}{\Gamma}_{cd}{}^{a}g^{cd}+\overset{m}{H}_{\kappa\lambda}{}^{a}g^{\kappa\lambda}\right)\left(\overset{m}{\Gamma}_{ef}{}^{b}g^{ef}+\overset{m}{H}_{\mu\nu}{}^{b}g^{\mu\nu}\right)+\frac{1}{2N^2}\overset{m}{\Gamma}_{cd}{}^{a}\overset{m}{\Gamma}_{ef}{}^{b}g^{ce}g^{df}$$
$$+\frac{1}{2N^2}\left(\overset{m}{H}{}^{\kappa\lambda a}\overset{m}{H}{}^{\mu\nu b}g_{\kappa\mu}g_{\lambda\nu}+2\overset{e}{H}{}^{ac\kappa}\overset{e}{H}{}^{bd\lambda}g_{cd}g_{\kappa\lambda}\right)$$
$$+\frac{1}{N^2}\bar{u}^a\bar{u}^b-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{\Gamma}_{cd}{}^{b)}g^{cd}-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{H}_{\kappa\lambda}{}^{b)}g^{\kappa\lambda}+o(N^{-2}),$$

we can finally write:

$$E_{u_0}[\tilde{u}^a\tilde{u}^b]= \tag{12}$$
$$=\frac{1}{N}g^{ab}-\frac{1}{N^2}g^{c(b}\partial_c\left(\overset{m}{\Gamma}_{de}{}^{a)}g^{de}\right)-\frac{1}{N^2}g^{c(b}\partial_c\left(\overset{m}{H}_{\kappa\lambda}{}^{a)}g^{\kappa\lambda}\right)+\frac{2}{N^2}g^{c(b}\partial_c\bar{u}^{a)}$$
$$+\frac{1}{4N^2}\left(\overset{m}{\Gamma}_{cd}{}^{a}g^{cd}+\overset{m}{H}_{\kappa\lambda}{}^{a}g^{\kappa\lambda}\right)\left(\overset{m}{\Gamma}_{ef}{}^{b}g^{ef}+\overset{m}{H}_{\mu\nu}{}^{b}g^{\mu\nu}\right)+\frac{1}{2N^2}\overset{m}{\Gamma}_{cd}{}^{a}\overset{m}{\Gamma}_{ef}{}^{b}g^{ce}g^{df}$$
$$+\frac{1}{2N^2}\left(\overset{m}{H}{}^{\kappa\lambda a}\overset{m}{H}{}^{\mu\nu b}g_{\kappa\mu}g_{\lambda\nu}+2\overset{e}{H}{}^{ac\kappa}\overset{e}{H}{}^{bd\lambda}g_{cd}g_{\kappa\lambda}\right)$$
$$+\frac{1}{N^2}\bar{u}^a\bar{u}^b-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{\Gamma}_{cd}{}^{b)}g^{cd}-\frac{1}{N^2}\bar{u}^{(a}\overset{m}{H}_{\kappa\lambda}{}^{b)}g^{\kappa\lambda}+o(N^{-2}).$$

Since $\hat{s}^I=O_p(N^{-1})$ and it is a smooth function of $\bar{x}$,

$$E_{u_0}[\hat{s}^I\hat{s}^J]=\frac{1}{N^2}\bar{s}^I\bar{s}^J+o(N^{-2}). \tag{13}$$

By (10) and relations (8),

$$E_{u_0}[\tilde{u}^a\tilde{u}^b\tilde{u}^c]= \tag{14}$$

$$\begin{aligned}
&= E_{u_0}\left[\left(g^{ad}\tilde{x}_d - \frac{1}{2N}\overset{m}{\Gamma}{}^{\alpha\beta a}\tilde{x}_\alpha\tilde{x}_\beta + \frac{1}{N}\bar{u}^a\right)\right. \\
&\quad \left.\left(g^{be}\tilde{x}_e - \frac{1}{2N}\overset{m}{\Gamma}{}^{\gamma\delta b}\tilde{x}_\gamma\tilde{x}_\delta + \frac{1}{N}\bar{u}^b\right)\left(g^{cf}\tilde{x}_f - \frac{1}{2N}\overset{m}{\Gamma}{}^{\epsilon\zeta c}\tilde{x}_\epsilon\tilde{x}_\zeta + \frac{1}{N}\bar{u}^c\right)\right] \\
&= \frac{1}{N^2}\left(T_{def}g^{ad}g^{be}g^{cf} - \frac{9}{2}\overset{m}{\Gamma}{}^{\alpha\beta(a}g^{b|e|}g^{c)f}g_{(\alpha\beta}g_{ef)} + 3g^{(ab}\bar{u}^{c)}\right) + o(N^{-2}) \\
&= \frac{1}{N^2}\left(T^{abc} - \frac{9}{2}\overset{m}{\Gamma}{}^{\alpha\beta(a}g^{b|e|}g^{c)f}g_{(\alpha\beta}g_{ef)} + 3g^{(ab}\bar{u}^{c)}\right) + o(N^{-2}),
\end{aligned}$$

$$E_{u_0}[\tilde{u}^a\tilde{u}^b\hat{s}^I] = \frac{1}{N^2}g^{ab}\bar{s}^I + o(N^{-2}) \tag{15}$$

and

$$E_{u_0}[\tilde{u}^a\tilde{u}^b\tilde{u}^c\tilde{u}^d] = \frac{3}{N^2}g^{(ab}g^{cd)} + o(N^{-2}). \tag{16}$$

We can now use (12), (13), (14), (15), (16) and (6) to calculate each term of expression (7). We have that:

$$\begin{aligned}
&\frac{1}{2}g_{ab}E_{u_0}(\tilde{u}^a\tilde{u}^b) = \\
&= \frac{m}{2N} - \frac{1}{2N^2}g^{ac}g^{de}\int\frac{\partial_a\partial_d\partial_e p\partial_c p}{p}\mu(dx) - \frac{1}{2N^2}g^{ac}g^{de}\int\frac{\partial_a\partial_c p\partial_d\partial_e p}{p}\mu(dx) \\
&\quad + \frac{1}{2N^2}g^{ac}g^{de}\int\frac{\partial_a\partial_c p\partial_d p\partial_e p}{p^2}\mu(dx) + \frac{5}{8N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{dec}g_{de}g^{ab} + \frac{1}{2N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{cde}g_{de}g^{ab} \\
&\quad - \frac{1}{2N^2}\overset{m}{\Gamma}_{abc}T^{dec}g_{de}g^{ab} + \frac{1}{N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{cab} - \frac{1}{2N^2}\overset{m}{\Gamma}_{abc}T^{abc} \\
&\quad - \frac{1}{2N^2}\partial_a\left(\overset{m}{H}_{\kappa\lambda}{}^a g^{\kappa\lambda}\right) + \frac{1}{N^2}\partial_a\bar{u}^a + \frac{1}{4N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{abc} \\
&\quad + \frac{1}{4N^2}\left[2\left(\overset{e}{H}{}^2_{\mathcal{P}}\right) + \left(\overset{m}{H}{}^2_{\mathcal{A}}\right)\right] + \frac{1}{4N^2}\overset{m}{\Gamma}_{abc}\overset{m}{H}{}^{\kappa\lambda c}g_{\kappa\lambda}g^{ab} - \frac{1}{2N^2}\overset{m}{\Gamma}_{cda}g^{cd}\bar{u}^a \\
&\quad + \frac{1}{2N^2}g_{ab}\left(\bar{u}^a - \frac{1}{2}\overset{m}{H}_{\kappa\lambda}{}^a g^{\kappa\lambda}\right)\left(\bar{u}^b - \frac{1}{2}\overset{m}{H}_{\mu\nu}{}^b g^{\mu\nu}\right) + o(N^{-2});
\end{aligned}$$

$$\frac{1}{2}g_{IJ}E_{u_0}[\hat{s}^I\hat{s}^J] = \frac{1}{2N^2}g_{IJ}\bar{s}^I\bar{s}^J + o(N^{-2});$$

$$\left(\frac{1}{2}\overset{\alpha}{\Gamma}_{abc} + \frac{\alpha}{3}T_{abc}\right)E_{u_0}[\tilde{u}^a\tilde{u}^b\tilde{u}^c] =$$

$$= -\frac{\alpha-5}{4N^2}\overset{m}{\Gamma}_{abc}\, T^{abc} - \frac{3\alpha-1}{8N^2}\overset{m}{\Gamma}_{abc}\, T^{dec}g_{de}g^{ab} + \frac{\alpha-3}{12N^2}T_{abc}T^{abc}$$
$$-\frac{1}{2N^2}\overset{\alpha}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{dea}g_{de}g^{bc} - \frac{1}{2N^2}\overset{-\alpha}{\Gamma}_{abc}\overset{m}{H}{}^{\kappa\lambda a}g_{\kappa\lambda}g^{bc} - \frac{1}{N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{cab}$$
$$-\frac{1}{4N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{dec}g_{de}g^{ab} - \frac{1}{4N^2}\overset{\alpha}{\Gamma}_{abc}\overset{m}{H}{}^{\kappa\lambda c}g_{\kappa\lambda}g^{ab} - \frac{1}{2N^2}\overset{m}{\Gamma}_{abc}\overset{m}{\Gamma}{}^{abc}$$
$$+\frac{1}{N^2}\overset{-\alpha}{\Gamma}{}_{ba}{}^{b}\bar{u}^a + \frac{1}{2N^2}\overset{\alpha}{\Gamma}_{abc}\, g^{ab}\bar{u}^c + o(N^{-2});$$

$$\left(\frac{1}{2}\overset{\alpha}{\Gamma}_{abI} + \overset{\alpha}{\Gamma}_{aIb} + \alpha T_{abI}\right) E_{u_0}[\tilde{u}^a\tilde{u}^b\hat{s}^I] = -\frac{1}{2N^2}\overset{\alpha}{H}_{abI}\, g^{ab}\bar{s}^I + o(N^{-2});$$

$$K_{abcd}E_{u_0}[\tilde{u}^a\tilde{u}^b\tilde{u}^c\tilde{u}^d] =$$
$$= \frac{(\alpha-3)(\alpha-5)}{32N^2}g^{ab}g^{cd}\int\frac{\partial_a p\partial_b p\partial_c p\partial_d p}{p^3}\mu(dx)$$
$$+\left(\frac{\alpha+1}{8N^2} - \frac{1}{2N^2}\right)g^{ab}g^{cd}\int\frac{\partial_a\partial_b p\partial_c p\partial_d p}{p^2}\mu(dx)$$
$$+\left(\frac{\alpha+1}{4N^2} - \frac{1}{N^2}\right)g^{ac}g^{bd}\int\frac{\partial_a\partial_b p\partial_c p\partial_d p}{p^2}\mu(dx)$$
$$+\frac{1}{2N^2}g^{ab}g^{cd}\int\frac{\partial_a\partial_b\partial_c p\partial_d p}{p}\mu(dx) + \frac{1}{8N^2}g^{ab}g^{cd}\int\frac{\partial_a\partial_b p\partial_c\partial_d p}{p}\mu(dx)$$
$$+\frac{1}{4N^2}g^{ac}g^{bd}\int\frac{\partial_a\partial_b p\partial_c\partial_d p}{p}\mu(dx) + o(N^{-2}).$$

Putting all together, with some further calculations, leads to the result. ∎

It should be noticed that each term in (5) is a scalar, that is, it does not depend on the coordinate system.

$$\frac{1}{4N^2}\left(\overset{m}{H}{}^2_{\mathcal{A}}\right) +$$
$$+\frac{1}{2N^2}g_{ab}\left(\bar{u}^a - \frac{1}{2}\overset{m}{H}{}_{\kappa\lambda}{}^{a}\, g^{\kappa\lambda}\right)\left(\bar{u}^b - \frac{1}{2}\overset{m}{H}{}_{\mu\nu}{}^{b}g^{\mu\nu}\right) + \frac{1}{N^2}\overset{\frac{1-\alpha}{2}}{\nabla}_a\left(\bar{u}^a - \frac{1}{2}\overset{m}{H}{}_{\kappa\lambda}{}^{a}g^{\kappa\lambda}\right)$$

is the only part involving the estimator $\hat{u}_N$;

$$\frac{1}{2N^2}\left(g_{IJ}\bar{s}^I\bar{s}^J - \overset{\alpha}{H}_{abI}\, g^{ab}\bar{s}^I\right)$$

is the only term depending on $\hat{s}$, and the rest depends only on the model. Expression (5), calculated in $s = 0$, can be used to study the asymptotic behaviour of $\hat{u}_N$ with respect to an α-divergence. For $\alpha = -1$ the last term disappears and we obtain the result present in Komaki (1995), Proposition 1, p.10.

Remark. From (5) we can obtain a decomposition of the average α-divergence from the true distribution to any predictive one, in two parts:

$$\begin{aligned} & E_{u_0}\left\{D_\alpha\left(p(x;u_0), p(x;\hat{u}_N(\bar{x}), \hat{s}(\bar{x}))\right)\right\} \\ = \; & E_{u_0}\left\{D_\alpha\left(p(x;u_0), p(x;\hat{u}_N(\bar{x}))\right)\right\} + \frac{1}{2N^2}\left[g_{IJ}\bar{s}^I\bar{s}^J - \overset{\alpha}{H}_{abI}\, g^{ab}\bar{s}^I\right] + o(N^{-2}). \end{aligned} \tag{17}$$

The first term in (17) depends on the choice of the estimative distribution and the other on the shift orthogonal to the model $\mathcal{P}$. It is well known that the problem of choosing a second-order efficient estimator $\hat{u}_N(\bar{x})$ has not, in general, a unique solution. On the other hand the following theorem solves the problem of the choice of the optimal shift orthogonal to the model.

Theorem 3.2 *The second order optimal choice of $\hat{s}^I(\bar{x})$ is given by:*

$$\hat{s}^I_{opt}(\bar{x}) = \frac{1}{2N}\, \overset{\alpha}{H}_{ab}{}^I(\hat{u}_N(\bar{x}))g^{ab}(\hat{u}_N(\bar{x})), \tag{18}$$

where $\hat{u}_N(\bar{x})$ is any asymptotically efficient estimator.

Proof: By Theorem 3.1,

$$\begin{aligned} & E_u\left\{D_\alpha\left(p(x;u), p(x;\hat{u}_N(\bar{x}), \hat{s}(\bar{x}))\right)\right\} - E_u\left\{D_\alpha\left(p(x;u), p(x;\hat{u}_N(\bar{x}), \hat{s}_{opt}(\bar{x}))\right)\right\} = \\ = \; & \frac{1}{2N^2} g_{IJ}\left(N^2 E(\hat{s}^I)E(\hat{s}^J) - \frac{1}{4}\, \overset{\alpha}{H}_{ab}{}^I\, \overset{\alpha}{H}_{cd}{}^J g^{ab} g^{cd}\right) \\ & - \frac{1}{2N^2}\, \overset{\alpha}{H}_{abI}\, g^{ab}\left(NE(\hat{s}^I) - \frac{1}{2}\, \overset{\alpha}{H}_{cd}{}^I g^{cd}\right) + o(N^{-2}) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2} g_{IJ} E(\hat{s}^I) E(\hat{s}^J) + \frac{1}{8N^2} \overset{\alpha}{H}_{ab}{}^{I} \overset{\alpha}{H}_{cd}{}^{J} g^{ab} g^{cd} g_{IJ} \\
&\quad - \frac{1}{2N} \overset{\alpha}{H}_{abI}\, g^{ab} E(\hat{s}^I) + o(N^{-2}) \\
&= \frac{1}{2N^2} g_{IJ} \left(N E(\hat{s}^I) - \frac{1}{2} \overset{\alpha}{H}_{ab}{}^{I} g^{ab} \right) \left(N E(\hat{s}^J) - \frac{1}{2} \overset{\alpha}{H}_{cd}{}^{J} g^{cd} \right) + o(N^{-2}).
\end{aligned}
$$

Since g_{IJ} is positive definite, $\hat{s}_{opt}$ is asymptotically optimal. ∎

Let us now define, for a, $b = 1, \ldots, m$,

$$
\begin{aligned}
h_{ab} &= \overset{\alpha}{\nabla}{}^{(\mathcal{F})}_{\partial_a l_\alpha} \partial_b l_\alpha - \overset{\alpha}{\nabla}_{\partial_a l_\alpha} \partial_b l_\alpha \qquad (19) \\
&= \partial_a \partial_b l_\alpha + \frac{1+\alpha}{2} p^{\frac{1-\alpha}{2}} g_{ab} - \overset{\alpha}{\Gamma}_{ab}{}^{c} \partial_c l_\alpha \\
&= p^{\frac{1-\alpha}{2}} \left(\partial_a \partial_b l + \frac{1-\alpha}{2} \partial_a l \partial_b l + \frac{1+\alpha}{2} g_{ab} - \overset{\alpha}{\Gamma}_{ab}{}^{c} \partial_c l \right).
\end{aligned}
$$

Vectors h_{ab} are, by definition, orthogonal to the original model $\mathcal{P}$. Moreover they belong to H_u. The following theorem explains the important role they play in our analysis.

Theorem 3.3 *The difference in average α-divergence from the true distribution, between the estimative distribution $p(x; \hat{u}_N(\bar{x}))$ and the optimal predictive distribution $p(x; \hat{u}_N(\bar{x}), \hat{s}_{opt}(\bar{x}))$, is maximal if and only if vectors h_{ab}, a, $b = 1, \ldots, m$, belong to the linear space spanned by the h_I's. In this case, the optimal predictive distribution is*

$$
\begin{aligned}
&p(x; \hat{u}, \hat{s}_{opt}) = \qquad (20) \\
&= p(x; \hat{u}) \left[1 + \frac{1}{2N} g^{ab} \left(\partial_a \partial_b l + \frac{1-\alpha}{2} \partial_a l \partial_b l + \frac{1+\alpha}{2} g_{ab} - \overset{\alpha}{\Gamma}_{ab}{}^{c} \partial_c l \right) \right] + o(N^{-1}).
\end{aligned}
$$

Proof: By (19) and definition of $\overset{\alpha}{H}_{abI}$, we have that

$$
\langle h_{ab}, h_I \rangle = \left\langle \partial_a \partial_b l_\alpha + \frac{1+\alpha}{2} p^{\frac{1-\alpha}{2}} g_{ab} - \overset{\alpha}{\Gamma}_{ab}{}^{c} \partial_c l_\alpha, h_I \right\rangle \qquad (21)
$$

$$= \langle \partial_a \partial_b l_\alpha, \partial_I l_\alpha \rangle = \overset{\alpha}{H}_{abI} \, .$$

By substituting (18) in (17),

$$\begin{aligned} &E_u \left\{ D_\alpha \left(p(x;u), p(x;\hat{u}_N(\bar{x})) \right) \right\} - E_u \left\{ D_\alpha \left(p(x;u), p(x;\hat{u}_N(\bar{x}), \hat{s}_{opt}(\bar{x})) \right) \right\} = \\ &= \frac{1}{8N^2} \overset{\alpha}{H}_{abI} \overset{\alpha}{H}_{cdJ} \, g^{ab} g^{cd} g^{IJ} + o(N^{-2}) = \frac{1}{8N^2} \| \overset{\alpha}{H}_{abI} \, g^{ab} g^{IJ} \partial_J l_\alpha \|^2 + o(N^{-2}) \\ &= \frac{1}{8N^2} \| \langle h_{ab}, h_I \rangle g^{ab} g^{IJ} h_J \|^2 + o(N^{-2}), \end{aligned}$$

that depends only on the projection of the h_{ab}'s on the linear space spanned by the h_I's. Thus, it is maximal if and only if the h_{ab}'s are included in this space and its maximal value is

$$\begin{aligned} &E_u \left\{ D_\alpha \left(p(x;u), p(x;\hat{u}_N(\bar{x})) \right) \right\} - E_u \left\{ D_\alpha \left(p(x;u), p(x;\hat{u}_N(\bar{x}), \hat{s}_{opt}(\bar{x})) \right) \right\} = \\ &\quad = \frac{1}{8N^2} \| \langle h_{ab}, h_I \rangle g^{ab} g^{IJ} h_J \|^2 + o(N^{-2}) = \frac{1}{8N^2} \| g^{ab} h_{ab} \|^2 + o(N^{-2}). \end{aligned} \tag{22}$$

In this situation, by (18), (21) and (19), we have that

$$\begin{aligned} \hat{s}^I_{opt} h_I &= \frac{1}{2N} \overset{\alpha}{H}_{ab}{}^I g^{ab} h_I = \frac{1}{2N} g^{ab} \langle h_{ab}, h_I \rangle g^{IJ} h_J = \frac{1}{2N} g^{ab} h_{ab} \\ &= \frac{1}{2N} p^{\frac{1-\alpha}{2}} g^{ab} \left(\partial_a \partial_b l + \frac{1-\alpha}{2} \partial_a l \partial_b l + \frac{1+\alpha}{2} g_{ab} - \overset{\alpha}{\Gamma}_{ab}{}^c \partial_c l \right), \end{aligned} \tag{23}$$

and the result follows by substituting (23) in (2). ∎

Remark. Including vectors h_{ab}'s on the enlarged model, allows us to attain the best improvement on the estimative distribution. For this reason, in the sequel we consider only models $\mathcal{M}$ containing directions h_{ab}'s. Since (20) depends only on the h_{ab}'s, Theorem 3.3 assures that the same optimal predictive distribution is obtained from any regular parametric model $\mathcal{M}$ containing $\mathcal{P}$ and the h_{ab}'s. In this sense, (20) gives a predictive distribution that can be considered optimal in the space $\mathcal{F}$ of all probability distributions equivalent to p.

In the case when $\mathcal{P}$ itself is a full exponential family, we have:

$$l(x;\theta) = \theta^i x_i - \psi(\theta),$$

$$\partial_a l(x;\theta) = x_a - \partial_a \psi(\theta)$$

and

$$\partial_a \partial_b l(x;\theta) = -\partial_a \partial_b \psi(\theta).$$

Since the second derivatives of $l(x;\theta)$ do not depend on x,

$$\partial_a \partial_b l = E(\partial_a \partial_b l) = -g_{ab}$$

and

$$\begin{aligned} \overset{\alpha}{\Gamma}_{abc} &= E\left[\left(\partial_a \partial_b l + \frac{1-\alpha}{2} \partial_a l \partial_b l\right) \partial_c l\right] \\ &= \frac{1-\alpha}{2} E(\partial_a l \partial_b l \partial_c l) = \frac{1-\alpha}{2} T_{abc}. \end{aligned}$$

Thus, we can write (20) in a simpler form:

$$\begin{aligned} & p(x;\hat{u},\hat{s}_{opt}) = \qquad (24) \\ & \quad = p(x;\hat{u}) \left[1 + \frac{1-\alpha}{4N} g^{ab} \left(\partial_a l \partial_b l - g_{ab} - T_{ab}^{\ \ c} \partial_c l\right)\right] + o(N^{-1}) \\ & \quad = p(x;\hat{u}) \left[1 + \frac{1-\alpha}{4N} \left(g^{ab}(x_a - \partial_a \psi)(x_b - \partial_b \psi) - m - g^{ab} T_{ab}^{\ \ c}(x_c - \partial_c \psi)\right)\right] \\ & \qquad + o(N^{-1}). \end{aligned}$$

Notice that for $\alpha = 1$ there is no correction, that is, we do not move out of the full exponential model. Moreover, for $\alpha = -1$ we obtain

$$\begin{aligned} & p(x;\hat{u},\hat{s}_{opt}) = \\ & \quad = p(x;\hat{u}) \left[1 + \frac{1}{2N} \left(g^{ab}(x_a - \partial_a \psi)(x_b - \partial_b \psi) - m - g^{ab} T_{ab}^{\ \ c}(x_c - \partial_c \psi)\right)\right] \\ & \qquad + o(N^{-1}), \end{aligned}$$

that is exactly the same result as Vidoni (1995), expression (3.1), p.7.

Example 3.4 We consider m-dimensional multivariate distributions $N(\mu, I_m)$,

$$p(x;\mu) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x^i - \mu^i)^2\right],$$

where $\mu = (\mu^i)$, $i = 1, \ldots, m$, is unknown. We have that

$$g_{ij}(\mu) = \delta_{ij}$$

and

$$\overset{\alpha}{\Gamma}_{ijk}(\mu) = 0,$$

for all α. Let now $x(l)$, $l = 1, \cdots, N$, be independent $N(\mu, I_m)$ and $\hat{\mu} = \hat{\mu}_N(\bar{x})$ be any estimator for the mean vector μ, where

$$\bar{x} = \frac{1}{N}\sum_{l=1}^{N} x(l).$$

By (19),

$$h_{ij} = \begin{cases} \frac{1-\alpha}{2} p^{\frac{1-\alpha}{2}}[(x^i - \mu^i)^2 - 1] & i = j \\ \frac{1-\alpha}{2} p^{\frac{1-\alpha}{2}}(x^i - \mu^i)(x^j - \mu^j) & i \neq j. \end{cases}$$

By (24),

$$p(x;\hat{\mu},\hat{s}_{opt}) = p(x;\hat{\mu})\left[1 + \frac{1-\alpha}{4N}\sum_{i=1}^{m}\left((x^i - \hat{\mu}^i)^2 - 1\right)\right] + o(N^{-1}).$$

Anyway, in this case we better substitute (23) in (3):

$$\begin{aligned} p(x;\hat{\mu},\hat{s}_{opt}) &= p(x;\hat{\mu})\exp\left[\frac{1}{2N} g^{ab}\frac{h_{ab}}{p^{\frac{1-\alpha}{2}}} - \phi_\mu\right] + o(N^{-1}) \\ &= p(x;\hat{\mu})\exp\left[\frac{1-\alpha}{4N}\sum_{i=1}^{m}\left((x^i - \hat{\mu}^i)^2 - 1\right) - \phi_\mu\right] + o(N^{-1}) \\ &= \sqrt{\frac{1 - \frac{1-\alpha}{2N}}{2\pi}} \exp\left[-\frac{1}{2}\left(1 - \frac{1-\alpha}{2N}\right)\sum_{i=1}^{m}(x^i - \hat{\mu}^i)^2\right] + o(N^{-1}). \end{aligned}$$

We thus have that the optimal predictive distribution is

$$N\left(\hat{\mu}, \left(1 - \frac{1-\alpha}{2N}\right)^{-1}\right).$$

For $\alpha = -1$, it is distributed as

$$N\left(\hat{\mu}, \left(1 - \frac{1}{N}\right)^{-1}\right)$$

that coincides, up to order N^{-1}, with the result of Barndorff-Nielsen and Cox (1994), p.318. From (22) we can calculate the difference in average α-divergence between the estimative distribution and the predictive distribution:

$$\begin{aligned}\frac{1}{8N^2}\|g^{ij}h_{ij}\|^2 &= \frac{(1-\alpha)^2}{32N^2}\left\|p^{\frac{1-\alpha}{2}}\sum_{i=1}^{m}\left[(x^i - \hat{\mu}^i)^2 - 1\right]\right\|^2 = \\ &= \frac{(1-\alpha)^2}{32N^2}\int\left[\sum_{i=1}^{m}\left[(x^i - \hat{\mu}^i)^2 - 1\right]\right]^2 p\,dx = \frac{(1-\alpha)^2}{16N^2}m,\end{aligned}$$

that does not depend on $\hat{\mu}$, the efficient estimator used. Let now $\hat{\mu}$ be the James-Stein estimator for μ, that is,

$$\hat{\mu}(\bar{x}) = \left(1 - \frac{m-2}{N\sum_{i=1}^{m}(\bar{x}^i)^2}\right)\bar{x}.$$

Then

$$\begin{aligned}\hat{\mu}_\infty(\bar{x}) &= \lim_{N\to\infty}\hat{\mu} = \bar{x}, \\ \bar{\mu}(\bar{x}) &= -\frac{m-2}{\sum_{i=1}^{m}(\bar{x}^i)^2}\bar{x}\end{aligned}$$

and

$$\bar{\mu} = \bar{\mu}(\mu) = -\frac{m-2}{\sum_{i=1}^{m}(\mu^i)^2}\mu.$$

We can use expression (5) with $s = 0$ to compare the two estimative distributions obtained respectively from the maximum likelihood estimator $\hat{\mu}_{mle} = \bar{x}$, and the James-Stein estimator:

$$\begin{aligned}&E_\mu\{D_\alpha(p(x;\mu), p(x;\hat{\mu}_{mle}))\} - E_\mu\{D_\alpha(p(x;\mu), p(x;\hat{\mu}))\} = \\ &\quad = -\frac{1}{2N^2}g_{ij}\bar{\mu}^i\bar{\mu}^j - \frac{1}{N^2}\partial_i\bar{\mu}^i + o(N^{-2}) = \frac{1}{2N^2}\frac{(m-2)^2}{\sum_{i=1}^{m}(\mu^i)^2} + o(N^{-2}).\end{aligned}$$

Remark. Let us consider an f-divergence D_f as loss function. Without loss of generality, we can suppose $f(1) = 0$ and $f''(1) = 1$. Theorem 3.1 can be easily generalized to this case by putting $\alpha = 2f'''(1) + 3$ and by substituting the coefficient

$$\frac{(\alpha - 11)(\alpha - 1)}{32}$$

of the term

$$\frac{Q_{abcd}g^{ab}g^{cd}}{N^2}$$

with

$$\beta = \frac{f^{(4)}(1) - 2f'''(1) - 4}{8}.$$

In fact, in the expantion of D_f, the first and second order terms remain unchanged. The coefficient of the third order term is

$$\frac{(f'''(1) + 3)}{6} T_{ABC} + \frac{1}{2} \overset{e}{\Gamma}_{ABC},$$

and it can be written as

$$\frac{\alpha}{3} T_{ABC} + \frac{1}{2} \overset{\alpha}{\Gamma}_{ABC}$$

with $\alpha = 2f'''(1) + 3$. The coefficient β is calculated by

$$\frac{f^{(4)}(1)}{8} - \frac{\alpha + 1}{8} = \frac{f^{(4)}(1) - 2f'''(1) - 4}{8}.$$

References

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

Amari, S. et al. (1987) *Differential Geometry in Statistical Inference*, Vol. 10 of *Lecture Notes-Monograph Series*, Institute of Mathematical Statistics, Hayward, chapter 2, pp. 19–94.

Barndorff-Nielsen, O. and Cox, D. (1994) *Inference and asymptotics*, Chapman and Hall, London.

Komaki, F. (1995) On asymptotic properties of predictive distributions, *Preprint series*, University of Tokyo.

Vidoni, P. (1995) A simple predictive density based on the p^*-formula, *Biometrika (to appear)*.

Relació dels últims Preprints publicats:

- **193** *Variability reduction with multiweighing proceedings.* Miquel Salicrú, Carlos González and Juan J. Barreiro. AMS Subject Classification: 62N10. December 1995.

- **194** *Some compactness results in real interpolation for families of Banach spaces.* María J. Carro and Jaak Peetre. 1980 Mathematics Subject Classification (1985 Revision): 46M37, 47B07. January 1996.

- **195** *Separatrix splitting for a high-frequency perturbation of the pendulum.* V. G. Gelfreich. 1991 Mathematics Subject Classification: 34C37, 34C29, 58F30, 70H05. January 1996.

- **196** *A second order Stratonovich differential equation with boundary conditions.* Aureli Alabert and David Nualart. AMS Subject Classification: 60H10, 60H07. January 1996.

- **197** *A Gentzen system equivalent to the BCK-logic.* Romà J. Adillon and Ventura Verdú. Mathematics Subject Classification: 03B20. January 1996.

- **198** *On Gentzen systems associated with the finite linear MV-algebras.* Àngel J. Gil, Antoni Torrens Torrell and Ventura Verdú. Mathematics Subject Classification: 03B22, 03F03, 03G20, 03G99. February 1996.

- **199** *Exponentially small splitting of separatrices under fast quasiperiodic forcing.* Amadeu Delshams, Vassili Gelfreich, Àngel Jorba and Tere M. Seara. AMS classification scheme numbers: 34C37, 58F27, 58F36, 11J25. February 1996.

- **200** *Existence and regularity of the density for solutions to stochastic differential equations with boundary conditions.* Arturo Kohatsu-Higa and Marta Sanz-Solé. AMS 1990 SCI: 60H07, 60H10, 60H99. March 1996.

- **201** *A forcing construction of thin-tall Boolean algebras.* Juan Carlos Martínez. 1991 Mathematics Subject Classification: 03E35, 06E99, 54612. March 1996.

- **202** *Weighted continuous metric scaling.* C. M. Cuadras and J. Fortiana. AMS Subject Classification: 62H25, 62G99. April 1996.

- **203** *Homoclinic orbits in the complex domain.* V.F. Lazutkin and C. Simó. AMS Subject Classification: 58F05. May 1996.

- **204** *Quasivarieties generated by simple MV-algebras.* Joan Gispert and Antoni Torrens. Mathematics Subject Classification: 03B50, 03G99, 06F99, 08C15. May 1996.

- **205** *Regularity of the law for a class of anticipating stochastic differential equations.* Carles Rovira and Marta Sanz-Solé. AMS Subject Classification: 60H07, 60H10. May 1996.

- **206** *Effective computations in Hamiltonian dynamics.* Carles Simó. AMS Subject Classification: 58F05, 70F07. May 1996.

- **207** *Small perturbations in a hyperbolic stochastic partial differential equation.* David Márquez–Carreras and Marta Sanz–Solé. AMS Subject Classification: 60H15, 60H07. May 1996.

- **208** *Transformed empirical processes and modified Kolmogorov-Smirnov tests for multivariate distributions.* A. Cabaña and E. M. Cabaña. AMS Subject Classification: Primary: 62G30, 62G20, 62G10. Secondary: 60G15. June 1996.

- **209** *Anticipating stochastic Volterra equations.* Elisa Alòs and David Nualart. AMS Subject Classification: 60H07, 60H20. June 1996.