Edgeworth Expansions of the Kullback-Leibler Information

Jen-Jen Lin* Naoki Saito † Richard A. Levine‡

submitted to JASA

November 25, 1999

Abstract

This paper proposes an approximation for the Kullback-Leibler information based on Edgeworth expansions. In information theory, entropy is a useful criterion for identifying a multivariate normal distribution. Comon (1994) proposed an Edgeworth-based expansion of neg-entropy in the univariate case. Based on the Edgeworth expansion of neg-entropy, a diagnosis is proposed here for checking multi-normality. Moreover, a measurement for Kullback-Leibler information is also proposed. We present numerical examples to demonstrate computational complexity and applications to diagnose multivariate normality, evaluate the differential entropy and choose the least statistically dependent basis from the wavelet packet dictionaries.

Keywords: Neg-entropy, differential entropy, cumulants, multivariate normal diagnostic, least statistically dependent basis, wavelet packet dictionary

1 Introduction

Given an m dimensional random vector $\mathbf{X}$ with density $p_{\mathbf{X}}$, the differential entropy $S(p_{\mathbf{X}})$, a measure of dispersion of the density $p_{\mathbf{X}}$, is defined by

$$S(p_{\mathbf{X}}) = -\int p_{\mathbf{X}}(\mathbf{u}) \log p_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u}\,. \tag{1}$$

Then the standardized neg-entropy is defined by

$$J(p_{\mathbf{X}}, \phi_{\mathbf{X}}) = S(\phi_{\mathbf{X}}) - S(p_{\mathbf{X}})\,, \tag{2}$$

*Corresponding author: Department of Statistics, Ming Chuan University, Taipei, Taiwan; email: jjlin@mcu.edu.tw
†Department of Mathematics, University of California, Davis, CA 95616, USA; email: saito@math.ucdavis.edu
‡Division of Statistics, University of California, Davis, CA 95616, USA; email: levine@wald.ucdavis.edu

This work is partially supported by grants from NSF-EPA DMS-99-78321 (JJL, NS, RAL) and NSF DMS-99-73032 (NS)

where $\phi_{\mathbf{X}}$ stands for the m-dimensional Gaussian density with the same mean and variance as $p_{\mathbf{X}}$. Neg-entropy may be written in another form, known as Kullback-Leibler information :

$$J(p_{\mathbf{X}}, \phi_{\mathbf{X}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{\phi_{\mathbf{x}}(\mathbf{u})} \, d\mathbf{u} \,. \tag{3}$$

Kullback-Leibler information, invariant under any invertible linear transformation, is a measure of "distance" in problems of discrimination. This measure can be built through density estimates of $p_{\mathbf{X}}$ (see Joe, 1989; Hall, 1987; and Hall and Morton, 1993). Density estimation, however, relies on the choice of kernel function and window size or bandwidth for each estimator. The computational and conceptual complexity in specifying these parameters limits the applicability of density estimation methods for estimating (3).

We propose an alternative method based on Edgeworth expansions to evaluate the Kullback-Leibler information. Comon (1994) and Jones and Sibson (1987) approximated the neg-entropy in one dimension by

$$J(p_Z, \phi_Z) = \frac{1}{12}\rho_3^2 + \frac{1}{48}\rho_4^2 + \frac{7}{48}\rho_3^4 - \frac{1}{8}\rho_3^2\rho_4 + o(n^{-2}) \,, \tag{4}$$

using an Edgeworth expansion. Here, ρ_r is the rth *standardized* cumulant of the standardized random variable Z, standardized sum of the random variables $X_1, \ldots, X_n$ with independent and identically distributon, and n is the number of available samples. The relationship between ρ_r and the cumulant κ_r of the random variable Z is

$$\rho_r = \kappa_r / \kappa_2^{r/2} \,.$$

We generalize this method towards an approximation of the neg-entropy for an m-dimensional random vector $\mathbf{X}$. In particular, the analogous Edgeworth expansion for neg-entropy of the standardized random vector $\mathbf{Z}$ is

$$J(p_{\mathbf{Z}}, \phi_{\mathbf{Z}}) = \frac{1}{12}\left[\sum_{i=1}^{m} (\kappa^{i,i,i})^2 + 3\sum_{i \neq j}^{m} (\kappa^{i,i,j})^2 + \frac{1}{6}\sum_{i \neq j \neq s}^{m} (\kappa^{i,j,s})^2\right] + O(n^{-\frac{3}{2}}) \,, \tag{5}$$

where the cumulant $\kappa^{i_1,\cdots,i_v}$ of $\mathbf{Z}$ is of order $n^{1-\frac{v}{2}}$. Moreover, the Edgeworth expansion can be applied to evaluate the Kullback-Leibler information of the m-vector $\mathbf{X}$

$$J(p_{\mathbf{X}}, q_{\mathbf{X}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{q_{\mathbf{x}}(\mathbf{u})} \, d\mathbf{u} \,. \tag{6}$$

Here, $p_{\mathbf{X}}$ and $q_{\mathbf{X}}$ must have the same first and second moments to apply the Edgeworth expansion. In one dimension, the KL distance is

$$J(p_Z, q_Z) = \frac{1}{12}(\rho_3 - \tilde{\rho}_3)^2 + O(n^{-\frac{3}{2}}) \,, \tag{7}$$

where ρ_r and $\tilde{\rho}_r$ denote the rth *standardized* cumulants of the random variable Z and $\tilde{Z}$ corresponding to p_Z and q_Z. While in m dimensions, the KL distance will be

$$J(p_{\mathbf{Z}}, q_{\mathbf{Z}}) = \frac{1}{12}\left[\sum_{i=1}^{m}(\kappa^{i,i,i} - \tilde{\kappa}^{i,i,i})^2 + 3\sum_{i\neq j}^{m}(\kappa^{i,i,j} - \tilde{\kappa}^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^{m}(\kappa^{i,j,s} - \tilde{\kappa}^{i,j,s})^2\right] + O(n^{-\frac{3}{2}}), \tag{8}$$

where the cumulants $\kappa^{i_1,\cdots,i_v}$ and $\tilde{\kappa}^{i_1,\cdots,i_v}$, corresponding to $p_{\mathbf{Z}}$ and $q_{\mathbf{Z}}$, are of order $n^{1-\frac{v}{2}}$.

For the general cases where $p_{\mathbf{x}}$ and $q_{\mathbf{x}}$ have different first and second moments, we also derive the formulas analogous to (7) and (8) in Section 7.

This paper elucidates two facts: First, the convergence rate of the corresponding Kullback-Leibler information based on the Edgeworth expansion is $O(n^{-3/2})$. On the other hand, the alternative density estimation approach to computing the Kullback-Leibler information can provide only root-n consistent estimators (Hall and Morton, 1993). Furthermore, the error rate of the histogram estimator not only depends on sample size n, but also on the choice of 'bin width' value h (Hall, 1987). The total error is, roughly, $O(h^2) + o(n^{-1/2})$. In the case of kernel estimation, the error is $o(n^{-1/2})$ when the dimension is less than (or equal to) 3; the estimator is much less sensitive to choices of the bandwidth h compared to the associated histogram estimator.

Second, the Kullback-Leibler information based on the Edgeworth expansion can be evaluated for any dimensional distribution as compared to density estimation (both histogram and kernel estimator) which can be performed only on low-dimensional distributions (1, 2, and 3 dimensions) in practice.

The paper is organized as follows. Section 2 is devoted to generalizing the Edgeworth expansion of the neg-entropy in m dimensions as shown in (5). Section 3 presents the derivation of the Edgeworth expansion of the generalized Kullback-Leibler information in both one and m dimensions displayed in (7) and (8). Section 4 discusses estimation of the generalized neg-entropy and Kullback-Leibler information via sample cumulants. In Section 5 we study the performance of our approximated distance measures through numerical examples. In Section 6, we illustrate our methods through three applications: diagnosing multivariate normality, evaluating neg-entropy, and compressing and decomposing an multidimensional image. In Section 7, we derive the general one-dimensional and m-dimensional Kullback-Leibler information where both densities have different first and second moments.

2 Neg-Entropy in m-dimensional space

We use the covariant and contravariant system (indexing random variables by lower and upper indices) to denote operations in high dimensional spaces (McCullagh, 1987). We present the definition of covariant-contravariant system and the corresponding properties of cumulants and covariant-contravariant Hermite polynomials in Appendix A.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent and identically distributed m-dimensional random vectors. Denote the components of each random vector by $\mathbf{X} = (X^1, \ldots, X^m)$, with mean $\mu = (\mu^1, \ldots, \mu^m)$ and moments

$$\kappa^{i_1 \ldots i_v} = E(X^{i_1} - \mu^{i_1}) \ldots (X^{i_v} - \mu^{i_v}) \,,$$

where $1 \leq i_k \leq m$, $1 \leq k \leq m$. Let $S_n = \sum_{i=1}^n \mathbf{X}_i$ and $\mathbf{Z} = (S_n - n\mu)/\sqrt{n}$ such that the cumulant $\kappa^{i_1, \ldots, i_v}$ of $\mathbf{Z}$ is of the order $n^{1-\frac{v}{2}}$. Then the Edgeworth expansion of $p_{\mathbf{Z}}$ up to order five about its best normal approximate is given by (Barndorff-Nielsen and Cox, 1989; Kendall and Stuart, 1977)

$$\begin{aligned} & p_{\mathbf{Z}}(\mathbf{z};\kappa) \\ = \;& \phi_m(\mathbf{z};\kappa) \quad \left[1 + \frac{1}{3!}\kappa^{i,j,k} h_{ijk}(\mathbf{z};\kappa) + \frac{1}{4!}\kappa^{i,j,k,l} h_{ijkl}(\mathbf{z};\kappa) + \frac{10}{6!}\kappa^{i,j,k}\kappa^{l,p,q} h_{ijklpq}(\mathbf{z};\kappa)\right] \\ & + \quad O(n^{-\frac{3}{2}}) \end{aligned} \tag{9}$$

where

$$\phi_m(\mathbf{z};\kappa) = (2\pi)^{-m/2} \{\det(\kappa)\}^{-1/2} \exp(-\frac{1}{2}\kappa_{i,j} z^i z^j) \,,$$

denotes the m-dimensional multivariate normal distribution of zero mean and covariance matrix $\kappa = [\kappa^{i,j}]$, with $\kappa^{i,j} = E(Z^i Z^j)$ and $[\kappa_{i,j}]$ represents κ^{-1}.

Using the covariant-contravariant system and the corresponding properties, we may simplify (9) by letting $v_1(\mathbf{z};\kappa)$, $v_2(\mathbf{z};\kappa)$, $v_3(\mathbf{z};\kappa)$ be the corresponding terms in the sum $\kappa^{i,j,k} h_{ijk}(\mathbf{z})$, $\kappa^{i,j,k,l} h_{ijkl}(\mathbf{z})$, and $\kappa^{i,j,k}\kappa^{l,p,q} h_{ijklpq}(\mathbf{z})$, with $v(\mathbf{z};\kappa) = v_1(\mathbf{z};\kappa) + v_2(\mathbf{z};\kappa) + v_3(\mathbf{z};\kappa)$. Note that there are m^2, m^3 and m^6 terms contained in the $v_1(\mathbf{z};\kappa)$, $v_2(\mathbf{z};\kappa)$, and $v_3(\mathbf{z};\kappa)$ respectively. Then (9) can be written as

$$p_{\mathbf{Z}}(\mathbf{z};\kappa) \;=\; \phi_m(\mathbf{z};\kappa)\; [1 + v(\mathbf{z};\kappa)] \;+ O(n^{-\frac{3}{2}}) \,.$$

Substituting this approximation into the neg-entropy defined in (3), we have the following expansion

$$\begin{aligned} J(p_{\mathbf{Z}}, \phi_{\mathbf{Z}}) &= \int p_{\mathbf{Z}}(\mathbf{z};\kappa) \log \frac{p_{\mathbf{Z}}(\mathbf{z};\kappa)}{\phi_{\mathbf{Z}}(\mathbf{z};\kappa)}\, d\mathbf{z} \\ &\approx \int \phi_m(\mathbf{z};\kappa)\; [1 + v(\mathbf{z};\kappa)] \log(1 + v(\mathbf{z};\kappa))\, d\mathbf{z} \;+ O(n^{-\frac{3}{2}}) \\ &\approx \int \phi_m(\mathbf{z};\kappa) \left[v(\mathbf{z};\kappa) + \frac{1}{2}(v(\mathbf{z};\kappa))^2\right] d\mathbf{z} \;+ O(n^{-\frac{3}{2}}) \\ &= \frac{1}{12}\left[\sum_{i=1}^m (\kappa^{i,i,i})^2 + 3\sum_{i\neq j}^m (\kappa^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^m (\kappa^{i,j,s})^2\right] \;+ O(n^{-\frac{3}{2}}) \end{aligned} \tag{10}$$

since

$$\int \phi_m(\mathbf{z};\kappa) v(\mathbf{z};\kappa)\, d\mathbf{z} \;=\; \int \phi_m(\mathbf{z};\kappa) v_1(\mathbf{z};\kappa)\, d\mathbf{z} + \int \phi_m(\mathbf{z};\kappa) v_2(\mathbf{z};\kappa)\, d\mathbf{z} + \int \phi_m(\mathbf{z};\kappa) v_3(\mathbf{z};\kappa)\, d\mathbf{z} \;=\; 0.$$

Note that there are m, $m(m-1)$ and $\frac{1}{6}m(m-1)(m-2)$ terms in each summation of (10).

In particular, the two-dimensional neg-entropy $J(p_{\mathbf{Z}}, \phi_{\mathbf{Z}})$ can be approximated by the Edgeworth expansion as follows (see Appendix B for details):

$$\begin{aligned} J(p_{\mathbf{Z}}, \phi_{\mathbf{Z}}) &= \int p_{\mathbf{Z}}(\mathbf{z};\kappa) \log \frac{p_{\mathbf{Z}}(\mathbf{z};\kappa)}{\phi_{\mathbf{Z}}(\mathbf{z};\kappa)}\, d\mathbf{z} \\ &\approx \int \phi_2(\mathbf{z};\kappa) \left[1 + \frac{1}{6}v_1(\mathbf{z};\kappa) + \frac{1}{24}v_2(\mathbf{z};\kappa) + \frac{10}{6!}v_3(\mathbf{z};\kappa)\right] \\ &\qquad \times \log\left(1 + \frac{1}{6}v_1(\mathbf{z};\kappa) + \frac{1}{24}v_2(\mathbf{z};\kappa) + \frac{10}{6!}v_3(\mathbf{z};\kappa)\right) d\mathbf{z} + O(n^{-\frac{3}{2}}) \\ &= \frac{1}{12}\left[(\kappa^{1,1,1})^2 + 3(\kappa^{1,1,2})^2 + 3(\kappa^{1,2,2})^2 + (\kappa^{2,2,2})^2\right] + O(n^{-\frac{3}{2}}). \end{aligned}$$

3 Kullback-Leibler information

The Kullback-Leibler information measure $J(p_{\mathbf{X}}, q_{\mathbf{X}})$, also called relative entropy or cross-entropy, is a measure of the 'distance' between two distributions $p_{\mathbf{X}}$ and $q_{\mathbf{X}}$ and is defined by

$$J(p_{\mathbf{X}}, q_{\mathbf{X}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{q_{\mathbf{x}}(\mathbf{u})}\, d\mathbf{u}\,. \tag{11}$$

The neg-entropy

$$J(p_{\mathbf{X}}, \phi_{\mathbf{X}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{\phi_{\mathbf{x}}(\mathbf{u})}\, d\mathbf{u} \tag{12}$$

is a special case of the Kullback-Leibler information. In Section 2, we derived the Edgeworth expansion of neg-entropy in one and m dimensions. Edgeworth expansions of the Kullback-Leibler information is analogous. In the case of one dimension, the Edgeworth expansion of q_Z up to order five about its best normal approximate is given by (Barndorff-Nielsen and Cox, 1989)

$$q_Z(z) = \phi_1(z)(1 + u(z)) + O(n^{-\frac{3}{2}})\,,$$

where

$$u(z) = \frac{1}{3!}\tilde{\rho}_3 H_3(z) + \frac{1}{4!}\tilde{\rho}_4 H_4(z) + \frac{10}{6!}\tilde{\rho}_3{}^2 H_6(z)\,,$$

and all other terms as defined in the previous sections. Substituting this expansion into the Kullback-Leibler information $J(p_Z, q_Z)$, we obtain

$$\begin{aligned} J(p_Z, q_Z) &= \int p_Z(z) \log \frac{p_Z(z)}{q_Z(z)}\, dz \\ &= \int p_Z(z) \log \frac{p_Z(z)}{\phi_1(z)(1+u(z))}\, dz \\ &= \int p_Z(z) \log \frac{p_Z(z)}{\phi_1(z)} - \int p_Z(z) \log(1+u(z))\, dz\,. \end{aligned} \tag{13}$$

The Edgeworth expansion of the first term, as shown by Comon (1994), is

$$J(p_Z, \phi_1) = \int p_Z(z) \log \frac{p_Z(z)}{\phi_1(z)}\, dz = \frac{1}{12}\rho_3^2 + O(n^{-\frac{3}{2}})\,.$$

To obtain the expansion of the second term of (13), the Edgeworth expansion of p_Z, similar to that of q_Z, is also needed:

$$p_Z(z) = \phi_1(z)(1 + v(z)) \; + O(n^{-\frac{3}{2}})\,,$$

where

$$v(z) = \frac{1}{3!}\rho_3 H_3(z) + \frac{1}{4!}\rho_4 H_4(z) + \frac{10}{6!}\rho_3^2 H_6(z)\,.$$

Then the second term of (13) can be expanded as

$$\begin{aligned}
& \int \phi_1(z)(1 + v(z)) \log(1 + u(z))\, dz \\
\approx \; & \int \phi_1(z)(1 + v(z))(u(z) - \frac{u^2(z)}{2} + O(n^{-\frac{3}{2}}))\, dz \\
= \; & \int \phi_1(z)(u(z) - \frac{u^2(z)}{2} + u(z)v(z))\, dz + O(n^{-\frac{3}{2}}) \\
= \; & -\frac{1}{12}\tilde{\rho_3}^2 + \frac{1}{6}\rho_3\tilde{\rho_3}
\end{aligned}$$

since

$$\begin{aligned}
\int \phi_1(z)u(z)\, dz &= 0 \\
-\frac{1}{2}\int \phi_1(z)u^2(z)\, dz &= -\frac{1}{12}\tilde{\rho_3}^2 \\
\int \phi_1(z)u(z)v(z)\, dz &= \frac{1}{6}\rho_3\, \tilde{\rho_3}\,.
\end{aligned}$$

Finally, the Edgeworth expansion of the Kullback-Leibler information is given by

$$J(p_Z, q_Z) = \frac{1}{12}\,(\rho_3 - \tilde{\rho_3})^2 \; + \; O(n^{-\frac{3}{2}})\,.$$

In the case of m dimensions, the Edgeworth expansion of $p_{\mathbf{Z}}$ and $q_{\mathbf{Z}}$ up to order five about their best normal approximate is given by (Barndorff-Nielsen and Cox, 1989)

$$\begin{aligned}
p_{\mathbf{Z}}(\mathbf{z};\kappa) &= \phi_m(\mathbf{z};\kappa)\; [1 + v(\mathbf{z};\kappa)] \; + O(n^{-\frac{3}{2}}) \\
q_{\mathbf{Z}}(\mathbf{z};\kappa) &= \phi_m(\mathbf{z};\kappa)\; [1 + u(\mathbf{z};\kappa)] \; + O(n^{-\frac{3}{2}})\,,
\end{aligned}$$

where $v(\mathbf{z};\kappa) = v_1(\mathbf{z};\kappa) + v_2(\mathbf{z};\kappa) + v_3(\mathbf{z};\kappa)$, $v_1(\mathbf{z};\kappa)$, $v_2(\mathbf{z};\kappa)$, $v_3(\mathbf{z};\kappa)$ are the corresponding terms in the summations over $\kappa^{i,j,k}h_{ijk}(\mathbf{z})$, $\kappa^{i,j,k,l}h_{ijkl}(\mathbf{z})$, and $\kappa^{i,j,k}\kappa^{l,p,q}h_{ijklpq}(\mathbf{z})$ respectively, and $u(\mathbf{z};\kappa) = u_1(\mathbf{z};\kappa) + u_2(\mathbf{z};\kappa) + u_3(\mathbf{z};\kappa)$, $u_1(\mathbf{z};\kappa)$, $u_2(\mathbf{z};\kappa)$, $u_3(\mathbf{z};\kappa)$ are the corresponding terms in the summations over $\tilde{\kappa}^{i,j,k}h_{ijk}(\mathbf{z})$, $\tilde{\kappa}^{i,j,k,l}h_{ijkl}(\mathbf{z})$, and $\tilde{\kappa}^{i,j,k}\tilde{\kappa}^{l,p,q}h_{ijklpq}(\mathbf{z})$, respectively.

Similarly to the one-dimensional case, the Kullback-Leibler information has the expansion

$$J(p_{\mathbf{Z}}, q_{\mathbf{Z}}) = \int p_{\mathbf{Z}}(\mathbf{z};\kappa) \log \frac{p_{\mathbf{Z}}(\mathbf{z};\kappa)}{\phi_{\mathbf{Z}}(\mathbf{z};\kappa)} - \int p_{\mathbf{Z}}(\mathbf{z};\kappa) \log(1 + u(\mathbf{z};\kappa))\, d\mathbf{z}\,. \tag{14}$$

The first term is the neg-entropy obtained in Section 2, namely

$$\begin{aligned} J(p_{\mathbf{Z}}, \phi_{\mathbf{Z}}) &= \int p_{\mathbf{Z}}(\mathbf{z};\kappa) \log \frac{p_{\mathbf{Z}}(\mathbf{z};\kappa)}{\phi_{\mathbf{Z}}(\mathbf{z};\kappa)}\, d\mathbf{z} \\ &\approx \frac{1}{12}\left[\sum_{i=1}^{m}(\kappa^{i,i,i})^2 + 3\sum_{i\neq j}^{m}(\kappa^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^{m}(\kappa^{i,j,s})^2\right] + O(n^{-\frac{3}{2}})\,. \end{aligned} \tag{15}$$

The second term can be expanded in an analogous way to the one-dimensional case

$$\begin{aligned} &\int \phi_{\mathbf{Z}}(\mathbf{z};\kappa)\,(1 + v(\mathbf{z};\kappa)) \log(1 + u(\mathbf{z};\kappa))\, d\mathbf{z} \\ \approx &\int \phi_{\mathbf{Z}}(\mathbf{z};\kappa)\,(1 + v(\mathbf{z};\kappa)) \left(u(\mathbf{z};\kappa) - \frac{u^2(\mathbf{z};\kappa)}{2} + O(n^{-\frac{3}{2}})\right) d\mathbf{z} \\ = &\int \phi_{\mathbf{Z}}(\mathbf{z};\kappa) \left(u(\mathbf{z};\kappa) - \frac{u^2(\mathbf{z};\kappa)}{2} + u(\mathbf{z};\kappa)v(\mathbf{z};\kappa)\right) d\mathbf{z} + O(n^{-\frac{3}{2}}) \\ = &-\frac{1}{12}\left[\sum_{i=1}^{m}(\tilde{\kappa}^{i,i,i})^2 + 3\sum_{i\neq j}^{m}(\tilde{\kappa}^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^{m}(\tilde{\kappa}^{i,j,s})^2\right] \\ &+\frac{1}{12}\left[2\sum_{i=1}^{m}\kappa^{i,i,i}\tilde{\kappa}^{i,i,i} + 6\sum_{i\neq j}^{m}\kappa^{i,i,j}\tilde{\kappa}^{i,i,j} + \frac{1}{3}\sum_{i\neq j\neq s}^{m}\kappa^{i,j,s}\tilde{\kappa}^{i,j,s}\right] + O(n^{-\frac{3}{2}})\,. \end{aligned}$$

Combining the first term (15) with the Edgeworth expansion of Kullback-Leibler information (14) we obtain

$$J(p_{\mathbf{Z}}, q_{\mathbf{Z}}) = \frac{1}{12}\left[\sum_{i=1}^{m}(\kappa^{i,i,i} - \tilde{\kappa}^{i,i,i})^2 + 3\sum_{i\neq j}^{m}(\kappa^{i,i,j} - \tilde{\kappa}^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^{m}(\kappa^{i,j,s} - \tilde{\kappa}^{i,j,s})^2\right] + O(n^{-\frac{3}{2}}). \tag{16}$$

4 Sample Cumulants

Both the Edgeworth expansion of the neg-entropy $J(p_{\mathbf{Z}}, \phi_{\mathbf{Z}})$ and the Kullback-Leibler information $J(p_{\mathbf{Z}}, q_{\mathbf{Z}})$ involve the third order cumulants $\kappa^{i,j,s}$ and $\tilde{\kappa}^{i,j,s}$ of the random vector $\mathbf{Z}$ and $\tilde{\mathbf{Z}}$, corresponding to $p_{\mathbf{Z}}$ and $q_{\mathbf{Z}}$ respectively, where

$$\begin{aligned} \kappa^{i,j,s} &= E(Z^i - \mu^i)(Z^j - \mu^j)(Z^s - \mu^s) \\ \tilde{\kappa}^{i,j,s} &= E(\tilde{Z}^i - \tilde{\mu}^i)(\tilde{Z}^j - \tilde{\mu}^j)(\tilde{Z}^s - \tilde{\mu}^s)\,. \end{aligned}$$

In the case of one dimension, the third order standardized cumulants ρ_3 and $\tilde{\rho}_3$ are needed. To apply all the approximations in Sections 2 and 3, we need to estimate all the third order cumulants.

The sample cumulants, the so-called k-statistics, are unbiased estimates of cumulants. Here, we use the notation introduced by McCullagh(1987). Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be independent and identically distributed m-dimensional random vectors where $\mathbf{X}_i$ has components $X_i^1, \cdots X_i^m$. For each cumulant of $\mathbf{X}_i$, κ, with appropriate superscripts, there is a unique polynomial symmetric function, denoted by k with matching superscripts, such that k is an unbiased estimate of κ. For example,

$$\begin{aligned} k^r &= n^{-1} \sum_{i=1}^{n} x_i^r \\ k^{r,t} &= \frac{1}{n} \phi^{ij} x_i^r x_j^t \\ k^{r,t,u} &= \frac{1}{n} \phi^{ijs} x_i^r x_j^t x_s^u \,. \end{aligned} \tag{17}$$

where

$$\phi^{ij} = \begin{cases} 1, & \text{if } i = j \quad ; \\ -\frac{1}{n-1}, & i \neq j \quad . \end{cases}$$

$$\phi^{ijs} = \begin{cases} 1, & \text{if } i = j = s \quad ; \\ -\frac{1}{n-1}, & i = j \neq s \quad ; \\ \frac{2}{(n-1)(n-2)}, & i \neq j \neq s \quad . \end{cases}$$

ensure the estimators are unbiased.

Another way to calculate the sample cumulants is to use the sample moments: $k^i = \frac{1}{n} \sum_{r=1}^{n} x_r^i$; $k^{ij} = \frac{1}{n} \sum_{r=1}^{n} x_r^i x_r^j$; $k^{ijs} = \frac{1}{n} \sum_{r=1}^{n} x_r^i x_r^j x_r^s$, and the relationship between cumulants and moments from (19). Then the third order cumulant can be expressed in terms of moments as

$$k^{i,j,s} = \frac{n^2}{(n-1)(n-2)} \left[k^{ijs} - k^i k^{js} - k^j k^{is} - k^s k^{ij} + 2\, k^i k^j k^s \right].$$

In this paper, we use the sample cumulants defined in (17). In the two dimensional case, there are four terms: $k^{1,1,1}$, $k^{2,2,2}$, $k^{1,1,2}$, and $k^{1,2,2}$. In the general case of m dimensions, there are m terms of $k^{i,i,i}$, $m(m-1)$ terms of $k^{i,i,j}$, and $\frac{m}{6}(m-1)(m-2)$ terms of $k^{i,j,s}$. See Anscombe (1961) for applications of k-statistics in detecting departures from the usual linear model assumption. Bickel (1978), Hinkley (1985), McCullagh and Pregibon (1987) and Brillinger (1994) have some more recent developments.

5 Numerical Examples

Sections 2 and 3 presented the expansions of four measures: the neg-entropy for one dimension and m dimensions; and the Kullback-Leibler information for one dimension and m dimensions. In this section, we illustrate the computation of these four approximations with random samples from the normal, t, Binomial, and Poisson distributions.

For the one-dimensional experiments, we use Normal $(0,1)$, t-distribution (df=1000), Binomial distribution ($N = 1000, p = 0.1$), and Poisson distribution (λ=1000). The true expected values of the neg-entropy of these distributions are all zero. The Edgeworth expansion of one-dimensional neg-entropy is

$$J(p_z, \phi_z) = \frac{1}{12}\rho_3^2 + O(n^{-\frac{3}{2}}) .$$

For the m-dimensional experiments, we simply use m-dimensional tensor products of the above four one-dimensional distributions, i.e., each m-dimensional distribution is a product of the same marginal distributions. The true expected values of the m-dimensional neg-entropy under these four distributions are again zero. The Edgeworth expansion of m-dimensional neg-entropy is

$$J(p_{\mathbf{z}}, \phi_{\mathbf{z}}) = \frac{1}{12}\left[\sum_{i=1}^{m}(\kappa^{i,i,i})^2 + 3\sum_{i\neq j}^{m}(\kappa^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^{m}(\kappa^{i,j,s})^2\right] + O(n^{-\frac{3}{2}}) .$$

As for the Kullback-Leibler information, we consider the following six cases for both one dimension and m dimensions: Normal$(0,I)$/Normal$(0,I)$, t(df=1000)/t(df=1000), Binomial($N = 1000, p = 0.1$)/Binomial($N = 1000, p = 0.1$), Poisson(λ=1000)/Poisson(λ=1000), Uniform$(1,100)$/Uniform$(1,100)$, and Normal $(0,I)/t$(df=1000). The expected values of the Kullback-Leibler information of the first five cases are exactly zero, and the last case is approximately zero. Recall that the approximate Kullback-Leibler information for the one-dimensional case is

$$J(p_z, q_z) = \frac{1}{12}(\rho_3 - \tilde{\rho}_3)^2 + O(n^{-\frac{3}{2}}) ,$$

and that for the m-dimensional case is

$$J(p_{\mathbf{z}}, q_{\mathbf{z}}) = \frac{1}{12}\left[\sum_{i=1}^{m}(k^{i,i,i} - \tilde{k}^{i,i,i})^2 + 3\sum_{i\neq j}^{m}(k^{i,i,j} - \tilde{k}^{i,i,j})^2 + \frac{1}{6}\sum_{i\neq j\neq s}^{m}(k^{i,j,s} - \tilde{k}^{i,j,s})^2\right] + O(n^{-\frac{3}{2}}).$$

5.1 Numerical Analysis and Simulation of Neg-entropy

Table 1 displays the results of the simulation studies on the Edgeworth expansion of the neg-entropy of four distributions mentioned above. For each of the four distributions, we consider $m = 1, 5, 8,$ and 10 dimensions. At each dimension of the four distributions, we evaluate the neg-entropy with 100 different sample sizes : $n = 50, 150, \cdots, 5000$. For each sample size, 100 simulations are conducted and the approximate neg-entropy is taken as the average over these 100 simulations. The expected error is $O(n^{-3/2})$ from (10), we thus have the expected convergence rate $O(n^{-1})$ in the expansion expression. To investigate the convergence rate $O(n^{-1})$, we consider the slope, log(neg-entropy) over $\log(n)$, among these 100 simulations. Theoretically, the slope should be -1, and as expected, our numerical results in Table 1 give slope values very close to -1 for all the cases.

distribution	$m = 1$	$m = 5$	$m = 8$	$m = 10$
Normal$(0, 1)$	0.00014 (-1.042)	0.00262 (-1.004)	0.00617 (-0.985)	0.00912 (-1.003)
t(1000)	0.00016 (-1.031)	0.00272 (-1.009)	0.00609 (-1.004)	0.01025 (-0.991)
Binomial(1000,0.1)	0.00018 (-1.014)	0.00283 (-1.035)	0.00643 (-1.005)	0.01067 (-0.987)
Poisson(1000)	0.00025 (-1.031)	0.00293 (-1.082)	0.00641 (-0.993)	0.01108 (-0.989)

Table 1: m-dim neg-entropy (slope) of four distributions

distributions	$m = 1$	$m = 5$	$m = 8$	$m = 10$
Normal$(0, I)$/Normal$(0, I)$	0.00023 (-1.009)	0.00445 (-1.016)	0.00921 (-1.055)	0.01369 (-0.993)
t(1000)/t(1000)	0.00047 (-1.008)	0.00458 (-0.999)	0.01021 (-1.012)	0.01523 (-0.995)
B(1000,0.1)/B(1000,0.1)	0.00044 (-1.011)	0.00481 (-1.073)	0.01221 (-1.057)	0.01372 (-1.042)
Poisson(1000)/Poisson(1000)	0.00082 (-1.019)	0.00538 (-1.054)	0.01232 (-1.034)	0.01361 (-0.984)
U(1,100)/U(1,100)	0.00093 (-1.005)	0.00237 (-1.032)	0.00544 (-1.061)	0.00987 (-0.991)
Normal$(0, I)$/t(1000)	0.00028 (-1.023)	0.00052 (-1.033)	0.01026 (-1.049)	0.01012 (-1.049)

Table 2: m-dim Kullback-Leibler information (slope) between m-dim distributions

5.2 Numerical Analysis and Simulation of Kullback-Leibler Information

Table 2 presents the results of simulation studies on the Kullback-Leibler information of the six cases mentioned in the beginning of this section. Similarly to the neg-entropy cases, we again consider $m = 1, 5, 8$, and 10. For each m, the approximate Kullback-Leibler information is evaluated for 100 different sample sizes $n = 50, 150, \cdots, 5000$. For each sample size, 100 simulations are conducted and the approximate Kullback-Leibler information is taken as the average over these 100 simulations. Again, to investigate the convergence rate $O(n^{-1})$, we consider the slope, log(neg-entropy) over $\log(n)$, among these 100 simulations. Theoretically, the slope should be -1, and as expected, our numerical results in Table 2 reach this slope value in all the cases.

Note that the Kullback-Leibler information measures the difference or entropy distance between two densities. It is known that the t-distribution can be approximated by normal distribution when degree of freedom tends to infinity. The sixth case of Table 2 shows the Kullback-Leibler information between the normal distribution and t-distribution with 1000 degrees of freedom. This Kullback-Leibler information tends to zero with the rate $O(n^{-1})$ as suggested by the slope measure.

6 Applications

The Kullback-Leibler information, as a distance measure, is commonly used in practice to differentiate between distributions. In this section, we apply the Kullback-Leibler information approximations presented to three problems: validating an assumed multivariate normal target population, computing the differential entropy, and choosing an approximate basis set for image decomposition.

6.1 Checking Multinormality

The first application is to validate a multivariate Gaussian distribution. Figure 1 demonstrates a two-dimensional non-Gaussian and Gaussian distributions. The Principal/Independent Component Analysis (PICA) algorithm (Lin et al., 1999) attempts to transform an image into a Gaussian distributed data. We wish to validate the performance of this algorithm using our neg-entropy computations. In particular, we test whether the PICA transformed data does in fact follow a bivariate Gaussian distribution. The empirical p-value of this test is computed as follows: Under the assumption of multivariate normality with the sample mean and the sample covariance matrix, we generate 100 samples from this multivariate normal distribution and evaluate the neg-entropy. We repeat this process 200 times to estimate the distribution of the neg-entropy. The empirical p-value is the proportion of the neg-entropies above the sample neg-entropy.

The PICA algorithm is an extension of the principal component analysis (PCA) and transforms a set of dependent random variables into approximately independent Gaussian random variables. There are forward and backward processes in the PICA algorithm. The PICA forward process, in its entirety, applies a nonlinear transformation to the PCA coordinates to obtain independent Gaussian coordinates. The transformation into marginally normal random variates follows from probability transformation and simulation theory. The PICA backward process has the advantage to sample correlated multivariate variables by sampling univariate independent Gaussian variables. We need only the forward PICA process in this example.

The left hand data in Figure 1 is a two-dimensional non-Gaussian original cigar data (Lin et al. 1999) to be transformed by the PICA algorithm. The right hand data is the two-dimensional Gaussian distribution which is the result of 5 iterations of the PICA algorithm applied to the cigar data. From the empirical p-value (0.001, 0.98) in Table 3, we can easily verify the transformed data does in fact follow a two-dimensional normal distribution.

We also illustrate diagnosing multivariate normality in a higher dimensional space. Table 4 presents the validation of a 15-dimensional transformation by PICA using the eye data of Lin et al. (1999). The results validate the performance of PICA (p-value = 1.00) in transforming the 15-dimensional non-Gaussian data into variates from a 15-dimensional Gaussian distribution.

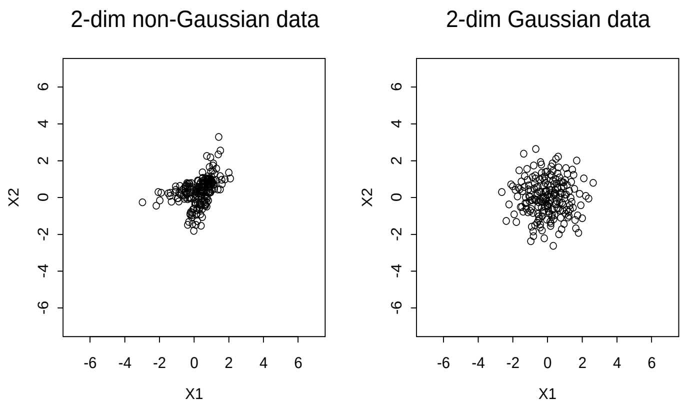

Figure 1: Plots of two-dimensional non-Gaussian and Gaussian distributions.

statistic	non-Gaussian distribution	Gaussian distribution
Neg-entropy	0.52	0.0013
emp. p-value	0.001	0.98

Table 3: Validation of multi-normality on the two-dimensional case

statistic	non-Gaussian distribution	Gaussian distribution
Neg-entropy	0.53	0.0034
emp. p-value	0.001	1.0

Table 4: Validation of multi-normality on the 15-dimensional case

n	via neg-entropy/abs.err (by Edgeworth expansion)	via differential entropy/abs.err (by density estimation)
100	1.439/0.019	1.389/0.031
200	1.425/0.005	1.392/0.028
300	1.424/0.004	1.404/0.016
400	1.424/0.004	1.413/0.007
500	1.422/0.002	1.425/0.005

Table 5: Numerical results of $S(\phi_X)$ (1.42 , theoretical value)

6.2 Evaluation of the differential entropy

In this section, we illustrate computation of the differential entropy using our Edgeworth approximation. We use this application to compare our approach to the estimated entropy of Hall and Morton (1993) based on density estimation. Note that the differential entropy

$$S(p_{\mathbf{X}}) = -\int p_{\mathbf{X}}(\mathbf{u}) \log p_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u} \tag{18}$$

may be written in terms of the neg-entropy

$$S(p_{\mathbf{X}}) = S(\phi_{\mathbf{X}}) - J(p_{\mathbf{X}}, \phi_{\mathbf{X}}) .$$

Table 5 shows the numerical results of $S(p_X)$ when $p_X = \phi_X$, (the theoretical value is 1.42) using neg-entropy and density estimation for samples of size $n = 100$, 200, 300, 400, and 500.

Table 6 presents the 2-dimensional numerical results of $S(p_X)$ when $p_X = \phi_X$ with three different covariances. Table 7 displays the 3-dimensional numerical results of $S(p_X)$ when $p_X = \phi_X$ with two different covariances. Table 8 shows the 4-dimensional, 5-dimensional and 8-dimensional numerical results of $S(p_Z)$ when $p_Z = \phi_Z$, Z denotes the standard normal random vectors. Here, in Table 5 – Table 8, we use the Edgeworth expansion with order $O(n^{-1.5})$. The convergence rate, as mentioned in Section 5, is $O(n^{-1})$ in these expansion expression . Recall that the density estimation approach is not applicable to populations of dimension larger than three.

The theoretical value of differential entropy for the p-dimensional normal random vector $\mathbf{Z}$ with covariance V is

$$S(\phi_{\mathbf{Z}}) = -\int \phi_{\mathbf{Z}}(\mathbf{z}) \log \phi_{\mathbf{Z}}(\mathbf{z})\, d\mathbf{z} = \frac{1}{2}\left[p + p\log(2\pi) + \log(\det(V))\right] .$$

As mentioned by Joe (1989) and Hall (1987), the method of density estimation is slow due to the choice of 'bandwidth' and kernel functions. Furthermore, density estimation is not applicable to evaluate differential entropy for dimensions of $\mathbf{X}$ greater than three. Of course, the Edgeworth

covariance	$cov = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$cov = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$	$cov = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$
true vale	2.8379	2.7907	2.3271
n	via neg-entropy/abs.err	via neg-entropy/abs.err	via neg-entropy/abs.err
100	2.5642/0.2737	2.6181/0.1726	2.4041/0.0769
200	2.7545/0.0833	2.6751/0.1156	2.2617/0.0653
300	2.7874/0.0505	2.8281/0.0374	2.2739/0.0530
400	2.8090/0.0288	2.7820/0.0086	2.3442/0.0172
500	2.8529/0.0149	2.7897/0.0009	2.3169/0.0101

Table 6: 2-dim numerical results of $S(\phi_X)$

covariance	$cov = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$cov = \begin{pmatrix} 1 & 0.8 & 0.6 \\ 0.8 & 1 & 0.4 \\ 0.6 & 0.4 & 1 \end{pmatrix}$
true vale	4.2568	3.5087
n	via neg-entropy/abs.err	via neg-entropy/abs.err
100	4.0904/0.1663	3.2448/0.2639
200	4.3208/0.0640	3.4428/0.0659
300	4.1976/0.0591	3.4757/0.0330
400	4.2800/0.0232	3.4737/0.0350
500	4.2730/0.0162	3.4899/0.0187

Table 7: 3-dim numerical results of $S(\phi_X)$

dimension	4-dim	5-dim	8-dim
true vale	5.6757	7.0946	11.35151
n	via neg-entropy/abs.err	via neg-entropy/abs.err	via neg-entropy/abs.err
100	5.3947/0.2810	6.7851/0.3095	11.0698/0.2816
200	5.4864/0.1893	7.3411/0.2464	11.1345/0.2169
300	5.5043/0.1714	6.9629/0.1317	11.1246/0.2268
400	5.5904/0.0852	6.9771/0.1175	11.1223/0.2291
500	5.7089/0.0331	7.0451/0.0495	11.2801/0.0713

Table 8: 4-dim, 5-dim, 8-dim numerical results of $S(\phi_{\mathbf{Z}})$ with identity covariance

expansion is not without drawbacks. If we want to include the high order cumulants (higher than fourth order), then the Edgeworth expansion will become more complicated and computationally complex, though we would attain a better order of approximation. Recall, though, that the order of approximation is $O(n^{-1})$, while the way via density estimation is $O(n^{-1/2})$ approximately. This difference in order explains the difference in absolute error between two techniques.

6.3 LSDB from the local basis dictionaries

Recent advances in imaging technology produce a large quantity of images over almost a continuous spatial spectrum as well as resolution. Image modeling is essential for the description and characterization of image features, large scale computations using images, and image compression. The most difficult problem in image modeling is the 'curse of dimensionality'. In particular, reliable estimates of probability density functions of high dimensional data, such as images, from a finite number of samples are hard to obtain in general. It is thus of paramount importance to extract relevant features from the images, reduce the dimensionality of the problem, and simplify the model by assuming statistical relationship among these features.

Image features are defined as the expansion coefficients of an image relative to some basis. Karhunen-Loève Basis is a decorrelated system. Saito (1994) developed and considered a local basis library to extract feature from image for classification and regression. The basis library consists of a collection of *local basis dictionaries* such as wavelet packets, local cosine/sine bases, or local Fourier bases. Each dictionary consists of a redundant number of the basis vectors with the specific characters in scale, position, and frequency. These basis vectors are organized as a quadtree in a hierarchical manner ranging from very localized spikes to global oscillations with different frequencies.

Image modeling techniques using the feature extractors have been proposed by various group of scientists. Saito (1998) developed an algorithm to find the *least statistically-dependent basis* (LSDB) by quickly selecting a basis from the local basis library mentioned above that is statistically independent coordinate system in the sense of relative entropy. He used the differential entropy $S(p_{\mathbf{X}})$ estimated by the method of density estimation as the selection criterion of LSDB:

$$B_{\text{LSDB}} = \arg\min_{B\in\mathcal{D}} \sum_{i=1}^{n} S(p_{X_i})\,.$$

Based on the relationship of differential entropy $S(p_{\mathbf{X}})$ and neg-entropy $J(p_{\mathbf{X}}, \phi_{\mathbf{X}})$

$$J(p_{\mathbf{X}}, \phi_{\mathbf{X}}) = S(\phi_{\mathbf{X}}) - S(p_{\mathbf{X}})\,,$$

we can rewrite the selection criterion as the form

$$B_{\text{LSDB}} = \arg\min_{B\in\mathcal{D}} \sum_{i=1}^{n} \left(S(\phi_{X_i}) - J(p_{X_i}, \phi_{X_i})\right)$$

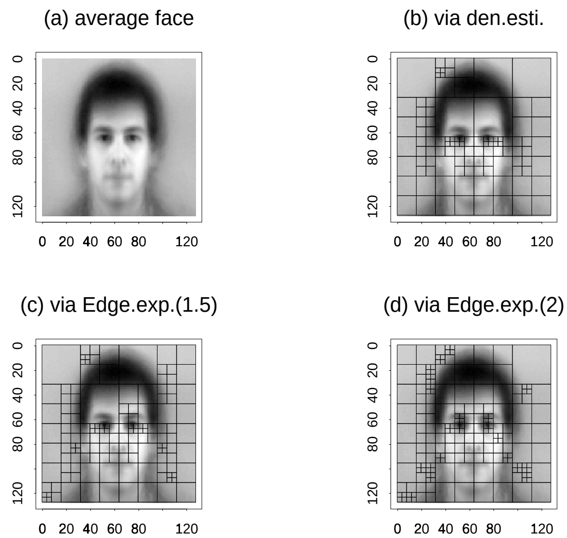

Figure 2: Comparison of LSDB chosen by using density estimation and Edgeworth Expansion with order $O(n^{-1.5})$ and $O(n^{-2})$

where neg-entropy $J(p_{X_i}, \phi_{X_i})$ can be estimated by the method of Edgeworth expansion.

To demonstrate the comparison between the LSDB selected by the method of density estimation and Edgeworth expansion, we use the data set of face images, 'Rogues' Gallery Problem'. This dataset consists of digitized pictures of faces of 143 people, provided by Prof. L. Sirovich at Brown University via Prof. M.V. Wickerhauser of Washington University. Among this dataset, we selected randomly 72 faces to be the training set. Each picture has the size of 128×128. Figure 2 (a) is the average face of the training set.

Figure 2 (b)–(d) show the partition patterns of LSDB selected from the local cosine dictionary by using density estimation and Edgeworth expansion with order $O(n^{-1.5})$ and $O(n^{-2})$. Here, there are 103 LSDB segments generated by the method of density estimation (Figure 2(b)); 142 LSDB segments are chosen by the method of Edgeworth expansion up to the order 1.5 ((Figure 2(c)); and 190 LSDB segments are chosen by the method of Edgeworth expansion up to the order 2 (Figure 2(d)). We observe that as the order of the Edgeworth expansion increases, the LSDB tries to split the image into finer segments. In particular, the LSDB segments in Figure 2(d) using the Edgeworth expansion up to the order 2, catches more information around the eye area than Figure 2(b), which was selected by density estimation.

7 General Kullback-Leibler information

In this section, we derive the Edgeworth expansion of the general Kullback-Leibler information ($p_{\mathbf{X}}$ and $q_{\mathbf{X}}$ do not share the same first and second moments). We present the expansion formula without the numerical results since the expansion is very complicated.

The Edgeworth expansion of the one-dimensional Kullback-Leibler information is

$$\begin{aligned} J(p_X, q_X) &= \int p_x(u) \log \frac{p_X(u)}{q_X(u)}\, du \\ &= \int p_X(u) \log \frac{p_X(u)}{\phi_{p_X}(u)}\, du + \int p_X(u) \log \frac{\phi_{p_X}(u)}{\phi_{q_X}(u)}\, du + \int p_X(u) \log \frac{\phi_{q_X}(u)}{q_X(u)}\, du \\ &= \frac{1}{12}\frac{\kappa_3^2}{\kappa_2^3} + \frac{1}{2}\left[\log \frac{\tilde{\kappa}_2}{\kappa_2} - 1 + \frac{1}{\tilde{\kappa}_2}(\kappa_1 - \tilde{\kappa}_1 + \kappa_2^{\frac{1}{2}})^2\right] - \left(\frac{\tilde{\kappa}_3 a_1}{6} + \frac{\tilde{\kappa}_4 a_2}{24} + \frac{\tilde{\kappa}_3^2 a_3}{72}\right) \\ &\quad -\frac{1}{2}\frac{\tilde{\kappa}_3^2}{36}\left[c_6 - \frac{6c_4}{\kappa_2} + \frac{9c_2}{\tilde{\kappa}_2^2}\right] - \frac{10\kappa_3\tilde{\kappa}_3(\kappa_1 - \tilde{\kappa}_1)(\kappa_2 - \tilde{\kappa}_2)}{\tilde{\kappa}_2^6} + O(n^{-\frac{3}{2}}). \end{aligned}$$

where

$$\begin{aligned} a_1 &= c_3 - \frac{3\alpha}{\tilde{\kappa}_2} \\ a_2 &= c_4 - \frac{6c_2}{\tilde{\kappa}_2} + \frac{3}{\tilde{\kappa}_2} \\ a_3 &= c_6 - \frac{15c_4}{\tilde{\kappa}_2} + \frac{45c_2}{\tilde{\kappa}_2^2} - \frac{15}{\tilde{\kappa}_2^3} \\ c_2 &= \alpha^2 + \beta^2 \\ c_3 &= \alpha^3 + 3\alpha\beta^2 \\ c_4 &= \alpha^4 + 6\alpha^2\beta^2 + 3\beta^4 \\ c_6 &= \alpha^6 + 15\alpha^4\beta^2 + 45\alpha^2\beta^4 + 15\beta^6 \\ \alpha &= \frac{\kappa_1 - \tilde{\kappa}_1}{\tilde{\kappa}_2} \\ \beta &= \frac{{\kappa_2}^{\frac{1}{2}}}{\tilde{\kappa}_2} \end{aligned}$$

Note that this formula reduces to (7) when $\kappa_1 = \tilde{\kappa_1} = 0$ and $\kappa_2 = \tilde{\kappa_2} = 1$.

For the high dimensional case, to avoid the complicated expansion, we derive the expansion up to order -1. The Edgeworth expansion of the m-dimensional Kullback-Leibler information is

$$\begin{aligned} J(p_{\mathbf{X}}, q_{\mathbf{X}}) &= \int p_{\mathbf{X}}(\mathbf{u}) \log \frac{p_{\mathbf{X}}(\mathbf{u})}{q_{\mathbf{X}}(\mathbf{u})}\, d\mathbf{u} \\ &= \int p_{\mathbf{X}}(\mathbf{u}) \log \frac{p_{\mathbf{X}}(\mathbf{u})}{\phi_{p_{\mathbf{X}}}(\mathbf{u})}\, d\mathbf{u} + \int p_{\mathbf{X}}(\mathbf{u}) \log \frac{\phi_{p_{\mathbf{X}}}(\mathbf{u})}{\phi_{q_{\mathbf{X}}}(\mathbf{u})}\, d\mathbf{u} + \int p_{\mathbf{X}}(\mathbf{u}) \log \frac{\phi_{q_{\mathbf{X}}}(\mathbf{u})}{q_{\mathbf{X}}(\mathbf{u})}\, d\mathbf{u} \\ &= \frac{1}{2}\left[\log \frac{\det(\tilde{\kappa})}{\det(\kappa)} - 1 + \sum_i \frac{(\kappa^{i,i})^2 + (\kappa^i - \tilde{\kappa}^i)^2}{(\tilde{\kappa}^{i,i})^2} + \sum_{i \neq j} \frac{2\tilde{\kappa}^{i,j}}{\tilde{\kappa}^{i,i}\tilde{\kappa}^{j,j}}(\kappa^{i,i}\kappa^{i,j} + (\kappa^i - \tilde{\kappa}^i)(\kappa^j - \tilde{\kappa}^j))\right] \end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{6}\sum_{i=1}^{m} \tilde{\kappa}^{i,i,i}\left[\sum_{J=1}^{m} \frac{(\kappa^{i,J})^2(\kappa^J-\tilde{\kappa}^J)(3\kappa_{i,i}+(\kappa_{i,J})^2-3\tilde{\kappa}_{i,i}(\kappa_{i,J})^2}{(\tilde{\kappa}^{i,J})^3}\right] \\
&- \frac{1}{6}\sum_{i\neq k} \tilde{\kappa}^{i,i,k}\left\{\sum_{J_1,J_2=1}^{m}\left[2\frac{\kappa^{k,J_2}\kappa^{i,J_1}(\kappa^{J_1}-\tilde{\kappa}^{J_1})}{\kappa^{i,k}\tilde{\kappa}^{k,J_2}(\tilde{\kappa}^{i,J_1})^2}+\frac{\kappa^{i,J_1}(\kappa^{J_2}-\tilde{\kappa}^{J_2})}{\tilde{\kappa}^{i,J_1}\tilde{\kappa}^{k,J_2}\kappa^{i,i}}+\frac{(\kappa^{J_1}-\tilde{\kappa}^{J_1})(\kappa^{J_2}-\tilde{\kappa}^{J_2})}{(\tilde{\kappa}^{i,J_1})^2\tilde{\kappa}^{k,J_2}}\right]\right\} \\
&- \frac{1}{6}\sum_{i\neq k} \tilde{\kappa}^{i,i,k}\left\{\sum_{J_1=1}^{m}\frac{2(\kappa^{J_1}-\tilde{\kappa}^{J_1})}{\tilde{\kappa}^{i,k}\tilde{\kappa}^{i,J_1}}+\sum_{J_2=1}^{m}\frac{\kappa^{J_2}-\tilde{\kappa}^{J_2}}{\tilde{\kappa}^{i,i}\tilde{\kappa}^{k,J_2}}\right\} \\
&- \frac{1}{6}\sum_{i\neq j\neq k} \tilde{\kappa}^{i,j,k}\left\{\sum_{J_1\neq J_2\neq J_3}\left[\frac{\kappa^{i,J_1}\kappa^{j,J_2}(\kappa^{J_3}-\tilde{\kappa}^{J_3})}{\kappa^{i,j}\tilde{\kappa}^{i,J_1}\tilde{\kappa}^{j,J_2}\tilde{\kappa}^{k,J_3}}+\frac{\kappa^{i,J_1}\kappa^{k,J_3}(\kappa^{J_2}-\tilde{\kappa}^{J_2})}{\kappa^{i,k}\tilde{\kappa}^{i,J_1}\tilde{\kappa}^{j,J_2}\tilde{\kappa}^{k,J_3}}+\frac{\kappa^{j,J_2}\kappa^{k,J_3}(\kappa^{J_1}-\tilde{\kappa}^{J_1})}{\kappa^{j,k}\tilde{\kappa}^{i,J_1}\tilde{\kappa}^{j,J_2}\tilde{\kappa}^{k,J_3}}\right]\right\} \\
&- \frac{1}{6}\sum_{i\neq j\neq k} \tilde{\kappa}^{i,j,k}\left\{\sum_{J_1\neq J_2\neq J_3}\frac{(\kappa^{J_1}-\tilde{\kappa}^{J_1})(\kappa^{J_2}-\tilde{\kappa}^{J_2})(\kappa^{J_3}-\tilde{\kappa}^{J_3})}{\tilde{\kappa}^{i,J_1}\tilde{\kappa}^{j,J_2}\tilde{\kappa}^{k,J_3}}\right\} \\
&+ \frac{1}{6}\sum_{i\neq j\neq k} \tilde{\kappa}^{i,j,k}\left\{\sum_{J_1=1}^{m}\frac{\kappa^{J_1}-\tilde{\kappa}^{J_1}}{\tilde{\kappa}^{i,J_1}\tilde{\kappa}^{j,k}}+\sum_{J_2=1}^{m}\frac{\kappa^{J_2}-\tilde{\kappa}^{J_2}}{\tilde{\kappa}^{j,J_2}\tilde{\kappa}^{i,k}}+\sum_{J_3=1}^{m}\frac{\kappa^{J_3}-\tilde{\kappa}^{J_3}}{\tilde{\kappa}^{k,J_3}\tilde{\kappa}^{i,j}}\right\} \\
&+ O(n^{-1}),
\end{aligned}
$$

where κ and $\tilde{\kappa}$ represent the covariance mtrices of $p_{\mathbf{X}}$ and $q_{\mathbf{X}}$.

These expressions are very complicated. We are currently conducting the numerical experiments for these general cases, which we hope to report at a later date.

Appendix A: covariant and contravariant system

To define the covariant and contravariant system more precisely, we start with a vector $\mathbf{x}$ with m components $x^1, x^2, \ldots, x^m$. We define u as a d-dimensional array whose elements are functions of the components of $\mathbf{x}$, taken d at a time. We write $u = u^{i_1 i_2 \cdots i_d} = (x^{i_1}, x^{i_2}, \ldots, x^{i_d})^T$ where the d components need not be distinct and T denotes the transposition. Consider the transformation $y = g(x)$ from $x^1, \ldots, x^m$ to new variables $y^1, \ldots, y^m$ and let $c_i^r \equiv c_i^r(x) = \frac{\partial y^r}{\partial x^i}$ having full rank for all x. If $\bar{u}$, the value of u for the transformed variables $y^r, r = 1, 2, \ldots, m$, satisfies

$$
\bar{u}^{r_1 r_2 \ldots r_d} = c_{i_1}^{r_1} c_{i_2}^{r_2} \cdots c_{i_d}^{r_d} u^{i_1 i_2 \ldots i_d}
$$

then u is said to be a *contravariant tensor*. On the other hand, if u is a *covariant tensor*, we write $u = u_{i_1 i_2 \cdots i_d}$ and the transformation law for covariant tensor is

$$
\bar{u}_{r_1 r_2 \ldots r_d} = d_{r_1}^{i_1} d_{r_2}^{i_2} \ldots d_{r_d}^{i_d} u_{i_1 i_2 \ldots i_d}
$$

where $d_r^i = \frac{\partial x^i}{\partial y^r}$, the matrix inverse of c_i^r, satisfies the relationship $c_i^r d_r^j = \delta_i^j = c_r^j d_i^r$.

Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be independent and identically distributed m-dimensional random vectors. Denote the components of each random vector by $\mathbf{X} = (X^1, \cdots, X^m)$, with mean $\mu = (\mu^1, \ldots, \mu^m)$

and moments

$$\kappa^{i_1 \dots i_v} = E(X^{i_1} - \mu^{i_1}) \dots (X^{i_v} - \mu^{i_v}) \,,$$

where $1 \le i_k \le m$, $1 \le k \le m$. The cumulants of $\mathbf{X}$ are the coefficients of the cumulant generating function

$$\kappa_{\mathbf{X}}(\mathbf{t}) = \log(M_{\mathbf{X}}(\mathbf{t})) = \sum_{i_1, \dots, i_v = 0}^{\infty} \frac{1}{i_1! \dots i_v!} \kappa^{i_1, \dots, i_v} t_{i_1} \dots t_{i_v} \,,$$

where $M_{\mathbf{X}}$ is the moment generating function of $\mathbf{X}$. Here, $\kappa^{i_1, \dots, i_v}$ is called the vth cumulant of $\mathbf{X}$. The following are the relationships between moments and cumulants.

$$\begin{aligned}
\kappa^{ij} &= \kappa^{i,j} + \kappa^i \kappa^j \\
\kappa^{ijk} &= \kappa^{i,j,k} + (\kappa^i \kappa^{j,k} + \kappa^j \kappa^{i,k} + \kappa^k \kappa^{i,j}) + \kappa^i \kappa^j \kappa^k \\
&= \kappa^{i,j,k} + \kappa^i \kappa^{j,k}[3] + \kappa^i \kappa^j \kappa^k \\
\kappa^{ijkl} &= \kappa^{i,j,k,l} + \kappa^i \kappa^{j,k,l}[4] + \kappa^{i,j} \kappa^{k,l}[3] + \kappa^i \kappa^j \kappa^{k,l}[6] + \kappa^i \kappa^j \kappa^k \kappa^l \,,
\end{aligned} \tag{19}$$

where $\kappa^i \kappa^{j,k}[3]$ is the sum over the three partitions of three indices. The following is a complete list of the 15 partitions of four items, one column for each of the five types (McCullagh, 1987)

$ijkl$	$i\|jkl$	$ij\|kl$	$i\|j\|kl$	$i\|j\|k\|l$
	$j\|ikl$	$ik\|jl$	$i\|k\|jl$	
	$k\|ijl$	$il\|jk$	$i\|l\|jk$	
	$l\|ijk$		$j\|k\|il$	
			$j\|l\|ik$	
			$k\|l\|ij$	

Let $S_n = \sum_{i=1}^{n} \mathbf{X}_i$ and $\mathbf{Z} = (S_n - \mu)/\sqrt{n}$ be the sum and standardized sum of random vectors $\mathbf{X}_1, \cdots, \mathbf{X}_n$ such that the cumulant $\kappa^{i_1, \dots, i_v}$ of $\mathbf{Z}$ is of the order $n^{1-\frac{v}{2}}$. Then the Edgeworth expansion of $p_{\mathbf{Z}}$ up to order five about its best normal approximate is given by (Barndorff-Nielsen and Cox, 1989; Kendall and Stuart, 1977)

$$\begin{aligned}
& p_{\mathbf{Z}}(\mathbf{z}; \kappa) \\
= \;& \phi_m(\mathbf{z}; \kappa) \left[1 + \frac{1}{3!} \kappa^{i,j,k} h_{ijk}(\mathbf{z}; \kappa) + \frac{1}{4!} \kappa^{i,j,k,l} h_{ijkl}(\mathbf{z}; \kappa) + \frac{10}{6!} \kappa^{i,j,k} \kappa^{l,p,q} h_{ijklpq}(\mathbf{z}; \kappa) \right] \\
+ \;& O(n^{-\frac{3}{2}})
\end{aligned} \tag{20}$$

where

$$\phi_m(\mathbf{z}; \kappa) = (2\pi)^{-m/2} \{\det(\kappa)\}^{-1/2} \exp(-\frac{1}{2} \kappa_{i,j} z^i z^j) \,,$$

denotes the m-dimensional multivariate normal distribution of zero mean and covariance matrix $\kappa = [\kappa^{i,j}]$, with $\kappa^{i,j} = E(Z^i Z^j)$ and $[k_{i,j}]$ represents κ^{-1}. The covariant Hermite polynomial $h_{i_1 \dots i_v}$ is defined as

$$\phi_m(\mathbf{x}; \kappa) \, h_{i_1 \dots i_v}(\mathbf{x}; \kappa) \; = \; (-1)^v \partial_{i_1} \dots \partial_{i_v} \, \phi_m(\mathbf{x}; \kappa) \,,$$

where $\partial_i = \partial/\partial x^i$. For the later use, the contravariant Hermite polynomial $h^{i_1 \cdots i_v}$ is defined as

$$\phi_m(\mathbf{x};\kappa)\, h^{i_1 \ldots i_v}(\mathbf{x};\kappa) \;=\; (-1)^v \partial^{i_1} \ldots \partial^{i_v}\, \phi_m(\mathbf{x};\kappa)\,,$$

with $\partial^i = \kappa^{i,j}\partial_j$. The first four covariant and contravariant Hermite polynomials are

$$\begin{aligned} h_i &= x_i\,, & h^i &= x^i\,,\\ h_{ij} &= x_i x_j - \kappa_{i,j}\,, & h^{ij} &= x^i x^j - \kappa^{i,j}\,,\\ h_{ijk} &= x_i x_j x_k - \kappa_{i,j} x_k[3]\,, & h^{ijk} &= x^i x^j x^k - \kappa^{i,j} x^k[3]\,,\\ h_{ijkl} &= x_i x_j x_k x_l - \kappa_{i,j} x_k x_l[6] + \kappa_{i,j}\kappa_{k,l}[3]\,, & h^{ijkl} &= x^i x^j x^k x^l - \kappa^{i,j} x^k x^l[6] + \kappa^{i,j}\kappa^{k,l}[3]\,, \end{aligned}$$

where the new notation x_i is defined as $x_i = \kappa_{i,j} x^j$.

Appendix B : Properties of Hermite polynomials

The expansion (9) may be simplified via certain properties of the Hermite polynomials (Skovgaard, 1981). First recall

$$h^{i_1 \ldots i_v}(\mathbf{x};\kappa) \;=\; \kappa^{i_1,j_1} \ldots \kappa^{i_v,j_v} h_{j_1 \ldots j_v}(\mathbf{x};\kappa)\,.$$

If the components of $\mathbf{X}$ are uncorrelated and of unit variance, then $\kappa^{i,i} = \kappa_{i,i} = 1$, $\kappa^{i,j} = \kappa_{i,j} = 0$. The covariant-contravariant Hermite polynomials for the multivariate distribution of $\mathbf{X}$ is then formed by taking all possible products of the Hermite polynomials (McCullagh, 1987):

$$\begin{aligned} h_{i\ldots i}(\mathbf{x}) &= h^{i\ldots i}(\mathbf{x}) = H_v(x^i), \quad i\ldots i \text{ denotes } v \text{ repetitions}\\ h_{i\ldots ij\ldots j}(\mathbf{x}) &= h^{i\ldots ij\ldots j}(\mathbf{x}) = H_{v-t}(x^i)H_t(x^j), \quad i\ldots i \text{ denotes } v-t \text{ repetitions}\\ & \qquad\qquad j\ldots j \text{ denotes } t \text{ repetitions}\\ h_{i_1\ldots i_v}(\mathbf{x}) &= h^{i_1\ldots i_v}(\mathbf{x}) = H_1(x^{i_1})\ldots H_1(x^{i_v}). \end{aligned}$$

Second, recall the useful orthogonality properties (Abramowitz and Stegun, 1972) in the Hermite polynomials

$$\begin{aligned} \int \phi(x) H_p(x) H_q(x)\, dx &= p!\,\delta_{pq}\\ \int \phi(x) H_3^2(x) H_4(x)\, dx &= 3!^3\,,\\ \int \phi(x) H_3^2(x) H_6(x)\, dx &= 6!\\ \int \phi(x) H_3^3(x)\, dx &= 0\\ \int \phi(x) H_3^4(x)\, dx &= 93 \cdot 3!^2\,, \end{aligned}$$

where $H_k(x)$ is the standard kth order Hermite polynomial. In the case of two dimensions ($m = 2$) with uncorrelated components of unit variance,

$$p_{\mathbf{Z}}(\mathbf{z};\kappa) \;=\; \phi_2(\mathbf{z};\kappa)\;[1 + v_1(\mathbf{z};\kappa) + v_2(\mathbf{z};\kappa) + v_3(\mathbf{z};\kappa)] \;+ O(n^{-\frac{3}{2}})$$

where

$$\begin{aligned}
v_1(\mathbf{z};\kappa) &= \kappa^{1,1,1}h_{111}(\mathbf{z}) + 3\kappa^{1,1,2}h_{112}(\mathbf{z}) + 3\kappa^{1,2,2}h_{122}(\mathbf{z}) + \kappa^{2,2,2}h_{222}(\mathbf{z})\,, \\
v_2(\mathbf{z};\kappa) &= \kappa^{1,1,1,1}h_{1111}(\mathbf{z}) + 4\kappa^{1,1,1,2}h_{1112}(\mathbf{z}) \\
&+ 6\kappa^{1,1,2,2}h_{1122}(\mathbf{z}) + 4\kappa^{1,2,2,2}h_{1222}(\mathbf{z}) + \kappa^{2,2,2,2}h_{2222}(\mathbf{z})\,, \\
v_3(\mathbf{z};\kappa) &= \kappa^{1,1,1}\kappa^{1,1,1}h_{111111}(\mathbf{z}) + 6\kappa^{1,1,1}\kappa^{1,1,2}h_{111112}(\mathbf{z}) \\
&+ 15\kappa^{1,1,1}\kappa^{1,2,2}h_{111122}(\mathbf{z}) + 20\kappa^{1,1,1}\kappa^{2,2,2}h_{111222}(\mathbf{z}) \\
&+ 15\kappa^{1,1,2}\kappa^{2,2,2}h_{112222}(\mathbf{z}) + 6\kappa^{1,2,2}\kappa^{2,2,2}h_{122222}(\mathbf{z}) \\
&+ \kappa^{2,2,2}\kappa^{2,2,2}h_{222222}(\mathbf{z})
\end{aligned}$$

and

$$\begin{aligned}
h_{111}(\mathbf{z}) = h^{111}(\mathbf{z}) &= H_3(z^1) & h_{112}(\mathbf{z}) &= h^{112}(\mathbf{z}) = H_2(z^1)H_1(z^2) \\
h_{122}(\mathbf{z}) = h^{122}(\mathbf{z}) &= H_1(z^1)H_2(z^2) & h_{222}(\mathbf{z}) &= h^{222}(\mathbf{z}) = H_3(z^2) \\
h_{1111}(\mathbf{z}) = h^{1111}(\mathbf{z}) &= H_4(z^1) & h_{1112}(\mathbf{z}) &= h^{1112}(\mathbf{z}) = H_3(z^1)H_1(z^2) \\
h_{1122}(\mathbf{z}) = h^{1122}(\mathbf{z}) &= H_2(z^1)H_2(z^2) & h_{1222}(\mathbf{z}) &= h^{1222}(\mathbf{z}) = H_1(z^1)H_3(z^2) \\
h_{111111}(\mathbf{z}) = h^{111111}(\mathbf{z}) &= H_6(z^1) & h_{111112}(\mathbf{z}) &= h^{111112}(\mathbf{z}) = H_5(z^1)H_1(z^2) \\
h_{111122}(\mathbf{z}) = h^{111122}(\mathbf{z}) &= H_4(z^1)H_2(z^2) & h_{111222}(\mathbf{z}) &= h^{111222}(\mathbf{z}) = H_3(z^1)H_3(z^2) \\
h_{112222}(\mathbf{z}) = h^{112222}(\mathbf{z}) &= H_2(z^1)H_4(z^2) & h_{122222}(\mathbf{z}) &= h^{122222}(\mathbf{z}) = H_1(z^1)H_5(z^2) \\
h_{222222}(\mathbf{z}) = h^{222222}(\mathbf{z}) &= H_6(z^2)\,.
\end{aligned}$$

The correlation term $v_1(\mathbf{z};\kappa)$, $v_2(\mathbf{z};\kappa)$, and $v_3(\mathbf{z};\kappa)$ will reduce to

$$\begin{aligned}
v_1(\mathbf{z};\kappa) &= \kappa^{1,1,1}H_3(z^1) + 3\kappa^{1,1,2}H_2(z^1)H_1(z^2) + 3\kappa^{1,2,2}H_1(z^1)H_2(z^2) + \kappa^{2,2,2}H_3(z^2)\,, \\
v_2(\mathbf{x};\kappa) &= \kappa^{1,1,1,1}H_4(z^1) + 4\kappa^{1,1,1,2}H_3(z^1)H_1(z^2) \\
&+ 6\kappa^{1,1,2,2}H_2(z^1)H_2(z^2) + 4\kappa^{1,2,2,2}H_1(z^1)H_3(z^2) + \kappa^{2,2,2,2}H_4(z^2)\,, \\
v_3(\mathbf{z};\kappa) &= \kappa^{1,1,1}\kappa^{1,1,1}H_6(z^1) + 6\kappa^{1,1,1}\kappa^{1,1,2}H_5(z^1)H_1(z^2) \\
&+ 15\kappa^{1,1,1}\kappa^{1,2,2}H_4(z^1)H_2(z^2) + 20\kappa^{1,1,1}\kappa^{2,2,2}H_3(z^1)H_3(z^2) \\
&+ 15\kappa^{1,1,2}\kappa^{2,2,2}H_2(z^1)H_4(z^2) + 6\kappa^{1,2,2}\kappa^{2,2,2}H_1(z^1)H_5(z^2) \\
&+ \kappa^{2,2,2}\kappa^{2,2,2}H_6(z^2)\,.
\end{aligned}$$

References

[1] Abramowitz, M and Stegun, I.A. (1972). *Handbook of Mathematical Functions*, Dover, New York.

[2] Anscombe, F.J. (1961). Examination of residuals. *Proc. 4th Berkeley Symp.*, **1**, 1-36.

[3] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

[4] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Inference and Asymptotics.* Chapman and Hall, London.

[5] Bickel, P.J. (1978). Using Residuals Robustly I : Tests for Heteroscedasticity Non-Linearity, *Ann. Statist.* **6**, 266-291.

[6] Brillinger, D.R. (1994). Some Basic Aspects and Uses of Higher-Order Spectra. *Signal Processing* **36**, 239-249.

[7] Comon, P. (1994). Independent Component Analysis, a New Concept? *Signal Processing* **36**, 287-314.

[8] Hall, P. (1987). On Kullback-Leibler Loss and Density Estimation. *The Annals of Statistics*, Vol. **15**, No. **4**, 1491-1519.

[9] Hall, P. and Morton, S.C. (1993). On the Estimation of Entropy. *Ann. Inst. Statist. Math.* , Vol. **45**, No. **1**, 69-88.

[10] Hinkley, D.V. (1985). Transformation Diagnostics for Linear Models. *Biometrika* **72**, 487-496.

[11] Joe, H. (1989). Estimation of Entropy and other Functionals of a Multivariate Density. *Ann. Inst. Statist. Math.* , Vol. **41**, 683-697.

[12] Jones, M.C. and Sibson, R. (1987). What is Projection Pursuit ?, *J.R.Statist.Soc.* A, **150**, Part I, 1-36.

[13] Kendall, M. and Stuart, A. (1977) *The Advanced Theory of Statistics*, Vol. **1**, New York.

[14] Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.

[15] Lin, J.J., Saito, N. and Levine, R.A. (1999). A New algorithm of ICA with application. In preparation.

[16] McCullagh, P. and Pregibon, D. (1987). K-statistics and Dispersion Effects in Regression. *Ann. Statist.*, Vol. **15**. No. **1**, 202-219.

[17] McCullagh, P. (1987). *Tensor Methods in Statistics.* Chapman And Hall, London.

[18] Saito, N. (1994). *Local Feature Extraction and Its Applications Using a Library of Bases.* PhD thesis, Department of Mathematics, Yale University, New Haven, CT 06520 USA.

[19] Saito, N. (1998). Least Statistically-Dependent Basis and its application to Image Modeling, In *Wavelet Applications in Signal and Image Processing VI*, Eds. A.F. Laine, M.A. Unser, and A. Aldroubi, Proc. SPIE 3458, 24-38.

[20] Skovgaard, I.M. (1981). Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions. *Scand. J. Statist.* **8**, 207-217.