

Article

# A Generic Formula and Some Special Cases for the Kullback–Leibler Divergence between Central Multivariate Cauchy Distributions

Nizar Bouhlel <sup>1,\*</sup>  and David Rousseau <sup>2</sup>

<sup>1</sup> ImhorPhen Unit, UMR INRAe IRHS, Institut Agro Rennes-Angers, Université d'Angers, 42 Rue Georges Morel, 49070 Beaucouzé, France

<sup>2</sup> LARIS, UMR INRAe IRHS, Université d'Angers, 62 Avenue Notre Dame du Lac, 49000 Angers, France; david.rousseau@univ-angers.fr

\* Correspondence: nizar.bouhlel@agrocampus-ouest.fr

**Abstract:** This paper introduces a closed-form expression for the Kullback–Leibler divergence (KLD) between two central multivariate Cauchy distributions (MCDs) which have been recently used in different signal and image processing applications where non-Gaussian models are needed. In this overview, the MCDs are surveyed and some new results and properties are derived and discussed for the KLD. In addition, the KLD for MCDs is showed to be written as a function of Lauricella D-hypergeometric series  $F_D^{(p)}$ . Finally, a comparison is made between the Monte Carlo sampling method to approximate the KLD and the numerical value of the closed-form expression of the latter. The approximation of the KLD by Monte Carlo sampling method are shown to converge to its theoretical value when the number of samples goes to the infinity.

**Keywords:** Multivariate Cauchy distribution (MCD); Kullback–Leibler divergence (KLD); multiple power series; Lauricella D-hypergeometric series



**Citation:** Bouhlel, N.; Rousseau, D. A Generic Formula and Some Special Cases for the Kullback–Leibler Divergence between Central Multivariate Cauchy Distributions. *Entropy* **2022**, *24*, 838. <https://doi.org/10.3390/e24060838>

Academic Editors: Karagrorgoriou Alexandros and Makrides Andreas

Received: 22 May 2022

Accepted: 15 June 2022

Published: 17 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Multivariate Cauchy distribution (MCD) belongs to the elliptical symmetric distributions [1] and is a special case of the multivariate  $t$ -distribution [2] and the multivariate stable distribution [3]. MCD has been recently used in several signal and image processing applications for which non-Gaussian models are needed. To name a few of them, in speckle denoizing, color image denoizing, watermarking, speech enhancement, among others. Sahu et al. in [4] presented a denoizing method for speckle noise removal applied to a retinal optical coherence tomography (OCT) image. The method was based on the wavelet transform where the sub-bands coefficients were modeled using a Cauchy distribution. In [5], a dual tree complex wavelet transform (DTCWT)-based despeckling algorithm was proposed for synthetic aperture radar (SAR) images, where the DTCWT coefficients in each subband were modeled with a multivariate Cauchy distribution. In [6], a new color image denoizing method in the contourlet domain was suggested for reducing noise in images corrupted by Gaussian noise where the contourlet subband coefficients were described by the heavy-tailed MCD. Sadreazami et al. in [7] put forward a novel multiplicative watermarking scheme in the contourlet domain where the watermark detector was based on the bivariate Cauchy distribution and designed to capture the across scale dependencies of the contourlet coefficients. Fontaine et al. in [8] proposed a semi-supervised multi-channel speech enhancement system where both speech and noise follow the heavy-tailed multi-variate complex Cauchy distribution.

Kullback–Leibler divergence (KLD), also called relative entropy, is one of the most fundamental and important measures in information theory and statistics [9,10]. KLD was first introduced and studied by Kullback and Leibler [11] and Kullback [12] to measure the

divergence between two probability mass functions in the case of discrete random variables and between two univariate or multivariate probability density functions in the case of continuous random variables. In the literature, numerous entropy and divergence measures have been suggested for measuring the similarity between probability distributions, such as Rényi [13] divergence, Sharma and Mittal [14] divergence, Bhattacharyya [15,16] divergence and Hellinger divergence measures [17]. Other general divergence families have been also introduced and studied like the  $\phi$ -divergence family of divergence measures defined simultaneously by Csiszár [18] and Ali and Silvey [19] where the KLD measure is a special case, the Bregman family divergence [20], the R-divergences introduced by Burbea and Rao [21–23], the statistical  $f$ -divergences [24,25] and recently the new family of a generalized divergence called the  $(h, \phi)$ -divergence measures introduced and studied in Menéndez et al. [26]. Readers are referred to [10] for details about these divergence family measures.

KLD has a specific interpretation in coding theory [27] and is therefore the most popular and widely used as well. Since information theoretic divergence and KLD in particular are ubiquitous in information sciences [28,29], it is therefore important to establish closed-form expressions of such divergence [30]. An analytical expression of the KLD between two univariate Cauchy distributions was presented in [31,32]. To date, the KLD of MCDs has no known explicit form, and it is in practice either estimated using expensive Monte Carlo stochastic integration or approximated. Monte Carlo sampling can efficiently estimate the KLD provided that a large number of independent and identically distributed samples is provided. Nevertheless, Monte Carlo integration is a too slow process to be useful in many applications. The main contribution of this paper is to derive a closed-form expression for the KLD between two central MCDs in a general case to benchmark future approaches while avoiding approximation using expensive Monte Carlo (MC) estimation techniques. The paper is organized as follows. Section 2 introduces the MCD and the KLD. Section 3 gives some definitions and propositions related to a multiple power series used to compute the closed-form expression of the KLD between two central MCDs. In Sections 4 and 5, expressions of some expectations related to the KLD are developed by exploiting the propositions presented in the previous section. Section 6 demonstrates some final results on the KLD computed for the central MCD. Section 7 presents some particular results such as the KLD for the univariate and the bivariate Cauchy distribution. Section 8 presents the implementation procedure of the KLD and a comparison with Monte Carlo sampling method. A summary and some conclusions are provided in the final section.

## 2. Multivariate Cauchy Distribution and Kullback–Leibler Divergence

Let  $\mathbf{X}$  be a random vector of  $\mathbb{R}^p$  which follows the MCD, characterized by the following probability density function (pdf) given as follows [2]

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = \frac{\Gamma(\frac{1+p}{2})}{\pi^{\frac{p}{2}} \Gamma(\frac{1}{2})} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \frac{1}{[1 + (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})]^{\frac{1+p}{2}}}. \quad (1)$$

This is for any  $\mathbf{x} \in \mathbb{R}^p$ , where  $p$  is the dimensionality of the sample space,  $\boldsymbol{\mu}$  is the location vector,  $\boldsymbol{\Sigma}$  is a symmetric, positive definite ( $p \times p$ ) scale matrix and  $\Gamma(\cdot)$  is the Gamma function. Let  $\mathbf{X}^1$  and  $\mathbf{X}^2$  be two random vectors that follow central MCDs with pdfs  $f_{\mathbf{X}^1}(\mathbf{x}|\boldsymbol{\Sigma}_1, p) = f_{\mathbf{X}^1}(\mathbf{x}|0, \boldsymbol{\Sigma}_1, p)$  and  $f_{\mathbf{X}^2}(\mathbf{x}|\boldsymbol{\Sigma}_2, p) = f_{\mathbf{X}^2}(\mathbf{x}|0, \boldsymbol{\Sigma}_2, p)$  given by (1). KLD provides an asymmetric measure of the similarity of the two pdfs. Indeed, the KLD between the two central MCDs is given by

$$\text{KL}(\mathbf{X}^1 || \mathbf{X}^2) = \int_{\mathbb{R}^p} \ln \left( \frac{f_{\mathbf{X}^1}(\mathbf{x}|\boldsymbol{\Sigma}_1, p)}{f_{\mathbf{X}^2}(\mathbf{x}|\boldsymbol{\Sigma}_2, p)} \right) f_{\mathbf{X}^1}(\mathbf{x}|\boldsymbol{\Sigma}_1, p) d\mathbf{x} \quad (2)$$

$$= E_{\mathbf{X}^1} \{ \ln f_{\mathbf{X}^1}(\mathbf{X}) \} - E_{\mathbf{X}^1} \{ \ln f_{\mathbf{X}^2}(\mathbf{X}) \}. \quad (3)$$

Since the KLD is the relative entropy defined as the difference between the cross-entropy and the entropy, we have the following relation:

$$KL(\mathbf{X}^1||\mathbf{X}^2) = H(f_{\mathbf{X}^1}, f_{\mathbf{X}^2}) - H(f_{\mathbf{X}^1}) \tag{4}$$

where  $H(f_{\mathbf{X}^1}, f_{\mathbf{X}^2}) = -E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^2}(\mathbf{X})\}$  denotes the cross-entropy and  $H(f_{\mathbf{X}^1}) = -E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^1}(\mathbf{X})\}$  the entropy. Therefore, the determination of KLD requires the expression of the entropy and the cross-entropy. It should be noted that the smaller  $KL(\mathbf{X}^1||\mathbf{X}^2)$ , the more similar are  $f_{\mathbf{X}^1}(x|\Sigma_1, p)$  and  $f_{\mathbf{X}^2}(x|\Sigma_2, p)$ . The symmetric KL similarity measure between  $\mathbf{X}^1$  and  $\mathbf{X}^2$  is  $d_{KL}(\mathbf{X}^1, \mathbf{X}^2) = KL(\mathbf{X}^1||\mathbf{X}^2) + KL(\mathbf{X}^2||\mathbf{X}^1)$ . In order to compute the KLD, we have to derive the analytical expressions of  $E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^1}(\mathbf{X})\}$  and  $E_{\mathbf{X}^1}\{\ln f_{\mathbf{X}^2}(\mathbf{X})\}$  which depend, respectively, on  $E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_1^{-1} \mathbf{X}]\}$  and  $E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\}$ . Consequently, the closed-form expression of the KLD between two zero-mean MCDs is given by

$$KL(\mathbf{X}^1||\mathbf{X}^2) = \frac{1}{2} \log \frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1+p}{2} \left( E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_1^{-1} \mathbf{X}]\} - E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\} \right). \tag{5}$$

To provide the expression of these two expectations, some tools based on the multiple power series are required. The next section presents some definitions and propositions used for this goal.

### 3. Definitions and Propositions

This section presents some definitions and exposes some propositions related to the multiple power series used to derive the closed-form expression of the expectation  $E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_1^{-1} \mathbf{X}]\}$  and  $E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\}$ , and as a consequence the KLD between two central MCDs.

**Definition 1.** The Humbert series of  $n$  variables, denoted  $\Phi_2^{(n)}$ , is defined for all  $x_i \in \mathbb{C}$ ,  $i = 1, \dots, n$ , by the following multiple power series (Section 1.4 in [33])

$$\Phi_2^{(n)}(b_1, \dots, b_n; c; x_1, \dots, x_n) = \sum_{m_1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \frac{(b_1)_{m_1} \dots (b_n)_{m_n}}{(c)_{\sum_{i=1}^n m_i}} \prod_{i=1}^n \frac{x_i^{m_i}}{m_i!}. \tag{6}$$

The Pochhammer symbol  $(q)_i$  indicates the  $i$ -th rising factorial of  $q$ , i.e., for an integer  $i > 0$

$$(q)_i = q(q+1) \dots (q+i-1) = \prod_{k=0}^{i-1} (q+k) = \frac{\Gamma(q+i)}{\Gamma(q)} \tag{7}$$

#### 3.1. Integral Representation for $\Phi_2^{(n)}$

**Proposition 1.** The following integral representation is true for  $Real\{c\} > Real\{\sum_{i=1}^n b_i\} > 0$  and  $Real\{b_i\} > 0$  where  $Real\{\cdot\}$  denotes the real part of the complex coefficients

$$\int \dots \int_{\Delta} \left(1 - \sum_{i=1}^n u_i\right)^{c - \sum_{i=1}^n b_i - 1} \prod_{i=1}^n u_i^{b_i - 1} e^{x_i u_i} du_i = B\left(b_1, \dots, b_n, c - \sum_{i=1}^n b_i\right) \Phi_2^{(n)}(b_1, \dots, b_n; c; x_1, \dots, x_n) \tag{8}$$

where  $\Delta = \{(u_1, \dots, u_n) | 0 \leq u_i \leq 1, i = 1, \dots, n; 0 \leq u_1 + \dots + u_n \leq 1\}$  and the multivariate beta function  $B$  is the extension of beta function to more than two arguments (called also Dirichlet function) defined as (Section 1.6.1 in [34])

$$B(b_1, \dots, b_n, b_{n+1}) = \frac{\prod_{i=1}^{n+1} \Gamma(b_i)}{\Gamma(\sum_{i=1}^{n+1} b_i)}. \tag{9}$$

**Proof.** The power series of exponential function is given by

$$e^{x_i u_i} = \sum_{m_i=0}^{\infty} \frac{x_i^{m_i}}{m_i!} u_i^{m_i}. \tag{10}$$

By substituting the expression of the exponential into the multiple integrals we have

$$\begin{aligned} & \int_{\Delta} \dots \int_{\Delta} \left(1 - \sum_{i=1}^n u_i\right)^{c - \sum_{i=1}^n b_i - 1} \prod_{i=1}^n u_i^{b_i - 1} e^{x_i u_i} du_i \\ &= \int_{\Delta} \dots \int_{\Delta} \left(1 - \sum_{i=1}^n u_i\right)^{c - \sum_{i=1}^n b_i - 1} \left(\prod_{i=1}^n \sum_{m_i=0}^{\infty} \frac{x_i^{m_i}}{m_i!} u_i^{m_i + b_i - 1} du_i\right) \\ &= \sum_{m_1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \left(\prod_{i=1}^n \frac{x_i^{m_i}}{m_i!}\right) \times \mathbf{I}_D \end{aligned} \tag{11}$$

where the multivariate integral  $\mathbf{I}_D$ , which is a generalization of a beta integral, is the type-1 Dirichlet integral (Section 1.6.1 in [34]) given by

$$\begin{aligned} \mathbf{I}_D &= \int_{\Delta} \dots \int_{\Delta} \left(1 - \sum_{i=1}^n u_i\right)^{c - \sum_{i=1}^n b_i - 1} \prod_{i=1}^n u_i^{b_i - 1} du_i \\ &= \frac{\prod_{i=1}^n \Gamma(b_i) \Gamma(c - \sum_{i=1}^n b_i)}{\Gamma(c + \sum_{i=1}^n b_i)}. \end{aligned} \tag{12}$$

Knowing that  $\Gamma(b_i + m_i) = \Gamma(b_i) (b_i)_{m_i}$ , the expression of  $\mathbf{I}_D$  can be written otherwise

$$\mathbf{I}_D = \frac{\prod_{i=1}^n \Gamma(b_i) \Gamma(c - \sum_{i=1}^n b_i)}{\Gamma(c)} \frac{\prod_{i=1}^n (b_i)_{m_i}}{(c)_{\sum_{i=1}^n m_i}}. \tag{13}$$

Finally, plugging (13) back into (12) leads to the final result

$$\frac{\Gamma(c - \sum_{i=1}^n b_i) \prod_{i=1}^n \Gamma(b_i)}{\Gamma(c)} \sum_{\substack{m_1, \dots, m_n \\ m_i=0}}^{+\infty} \frac{\prod_{i=1}^n (b_i)_{m_i}}{(c)_{\sum_{i=1}^n m_i}} \prod_{i=1}^n \frac{x_i^{m_i}}{m_i!} = B\left(b_1, \dots, b_n, c - \sum_{i=1}^n b_i\right) \Phi_2^{(n)}(b_1, \dots, b_n; c; x_1, \dots, x_n) \tag{14}$$

□

Given Proposition 1, we consider the particular cases  $n = \{1, 2\}$  one by one as follows:  
Case  $n = 1$

$$\frac{1}{B(b_1, c - b_1)} \int_0^1 u_1^{b_1 - 1} e^{x_1 u_1} (1 - u_1)^{c - b_1 - 1} du_1 = \sum_{m_1=0}^{\infty} \frac{(b_1)_{m_1}}{(c)_{m_1}} \frac{x_1^{m_1}}{m_1!} = \Phi_2^{(1)}(b_1; c; x_1) = {}_1F_1(b_1, c; x_1) \tag{15}$$

where  ${}_1F_1(\cdot)$  is the confluent hypergeometric function of the first kind (Section 9.21 in [35]).

Case  $n = 2$

$$\begin{aligned} & \frac{1}{B(b_1, b_2, c - b_1 - b_2)} \iint_{\substack{u_1 \geq 0, u_2 \geq 0, \\ u_1 + u_2 \leq 1}} u_1^{b_1 - 1} u_2^{b_2 - 1} e^{x_1 u_1 + x_2 u_2} (1 - u_1 - u_2)^{c - b_1 - b_2 - 1} du_1 du_2 \\ &= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \frac{(b_1)_{m_1} (b_2)_{m_2}}{(c)_{m_1 + m_2}} \frac{x_1^{m_1}}{m_1!} \frac{x_2^{m_2}}{m_2!} = \Phi_2^{(2)}(b_1, b_2; c; x_1, x_2) = \Phi_2(b_1, b_2, c; x_1, x_2) \end{aligned} \tag{16}$$

where the double series  $\Phi_2$  is one of the components of the Humbert series of two variables [36] that generalize Kummer’s confluent hypergeometric series  ${}_1F_1$  of one variable. The double series  $\Phi_2$  converges absolutely at any  $x_1, x_2 \in \mathbb{C}$ .

### 3.2. Multiple Power Series $F_N^{(n)}$

**Definition 2.** We define a new multiple power series, denoted by  $F_N^{(n)}$  and given by

$$F_N^{(n)}(a; b_1, \dots, b_n; c, c_n; x_1, \dots, x_n) = x_n^{-a} \sum_{\substack{m_1, \dots, \\ m_n=0}}^{+\infty} \frac{(a)_{\sum_{i=1}^n m_i} (a - c_n + 1)_{\sum_{i=1}^n m_i} \prod_{i=1}^{n-1} (b_i)_{m_i}}{(a + b_n - c_n + 1)_{\sum_{i=1}^n m_i} (c)_{\sum_{i=1}^{n-1} m_i} \prod_{i=1}^{n-1} \left(\frac{x_i}{x_n}\right)^{m_i}} \frac{1}{m_i!} \frac{(1 - x_n^{-1})^{m_n}}{m_n!}. \tag{17}$$

The multiple power series (17) is absolutely convergent on the region  $|x_i x_n^{-1}| + |1 - x_n^{-1}| < 1$  in  $\mathbb{C}^n, \forall i \in \{1, \dots, n - 1\}$ .

The multiple power series  $F_N^{(n)}(\cdot)$  can also be transformed into two other expressions as follows

$$F_N^{(n)}(a; b_1, \dots, b_n; c, c_n; x_1, \dots, x_n) = \sum_{\substack{m_1, \dots, \\ m_n=0}}^{+\infty} \frac{(a - c_n + 1)_{\sum_{i=1}^{n-1} m_i} (b_n)_{m_n} (a)_{\sum_{i=1}^n m_i} \prod_{i=1}^{n-1} (b_i)_{m_i}}{(a + b_n - c_n + 1)_{\sum_{i=1}^n m_i} (c)_{\sum_{i=1}^{n-1} m_i} \prod_{i=1}^{n-1} m_i!} \frac{x_i^{m_i} (1 - x_n)^{m_n}}{m_n!}, \tag{18}$$

$$= x_n^{1-c_n} \sum_{\substack{m_1, \dots, \\ m_n=0}}^{+\infty} \frac{(a - c_n + 1)_{\sum_{i=1}^n m_i} (b_n - c_n + 1)_{m_n} (a)_{\sum_{i=1}^{n-1} m_i} \prod_{i=1}^{n-1} (b_i)_{m_i}}{(a + b_n - c_n + 1)_{\sum_{i=1}^n m_i} (c)_{\sum_{i=1}^{n-1} m_i} \prod_{i=1}^{n-1} m_i!} \frac{x_i^{m_i} (1 - x_n)^{m_n}}{m_n!}. \tag{19}$$

By Horn’s rule for the determination of the convergence region (see [37], Section 5.7.2), the multiple power series (18) and (19) are absolutely convergent on region  $|x_i| < 1, \forall i \in \{1, \dots, n - 1\}, |1 - x_n| < 1$  in  $\mathbb{C}^n$ .

Equation (18) can then be deduced from (17) by using the following development where the  $F_N^{(p)}$  function can be written as

$$F_N^{(n)}(a; b_1, \dots, b_n; c, c_n; x_1, \dots, x_n) = x_n^{-a} \sum_{\substack{m_1, \dots, \\ m_{n-1}=0}}^{+\infty} \frac{(a)_{\sum_{i=1}^{n-1} m_i} (a - c_n + 1)_{\sum_{i=1}^{n-1} m_i} \prod_{i=1}^{n-1} (b_i)_{m_i}}{(a + b_n - c_n + 1)_{\sum_{i=1}^{n-1} m_i} (c)_{\sum_{i=1}^{n-1} m_i}} \times \prod_{i=1}^{n-1} \left(\frac{x_i}{x_n}\right)^{m_i} \frac{1}{m_i!} \sum_{m_n=0}^{\infty} \frac{(\alpha)_{m_n} (\alpha - c_n + 1)_{m_n} (1 - x_n^{-1})^{m_n}}{(\alpha + b_n - c_n + 1)_{m_n} m_n!} \tag{20}$$

and  $\alpha = a + \sum_{i=1}^{n-1} m_i$  is used here to alleviate writing equations. Using the definition of Gauss’ hypergeometric series  ${}_2F_1(\cdot)$  [34] and the Pfaff transformation [38], we can write

$$\sum_{m_n=0}^{\infty} \frac{(\alpha)_{m_n} (\alpha - c_n + 1)_{m_n} (1 - x_n^{-1})^{m_n}}{(\alpha + b_n - c_n + 1)_{m_n} m_n!} = {}_2F_1(\alpha, \alpha - c_n + 1; \alpha + b_n - c_n + 1; 1 - x_n^{-1}) \tag{21}$$

$$= x_n^\alpha {}_2F_1\left(\alpha, b_n; \alpha + b_n - c_n + 1; 1 - x_n\right) \tag{22}$$

$$= x_n^\alpha \sum_{m_p=0}^{\infty} \frac{(\alpha)_{m_n} (b_n)_{m_n}}{(\alpha + b_n - c_n + 1)_{m_n} m_n!} (1 - x_n)^{m_n}. \tag{23}$$

By substituting (23) into (20), and using the following two relations:

$$(a)_{\sum_{i=1}^{n-1} m_i} (\alpha)_{m_n} = (a)_{\sum_{i=1}^n m_i}, \tag{24}$$

$$(a + b_n - c_n + 1)_{\sum_{i=1}^{n-1} m_i} (\alpha + b_n - c_n + 1)_{m_n} = (a + b_n - c_n + 1)_{\sum_{i=1}^n m_i} \tag{25}$$

we can get (18).

The second transformation is given as follows

$$\begin{aligned}
 & {}_2F_1(\alpha, \alpha - c_n + 1; b_n - c_n + \alpha + 1; 1 - x_n^{-1}) \\
 &= x_n^{\alpha - c_n + 1} {}_2F_1(b_n - c_n + 1, \alpha - c_n + 1; \alpha + b_n - c_n + 1; 1 - x_n) \tag{26}
 \end{aligned}$$

$$= x_n^{\alpha - c_n + 1} \sum_{m_n=0}^{\infty} \frac{(\alpha - c_n + 1)_{m_n} (b_n - c_n + 1)_{m_n} (1 - x_n)^{m_n}}{(\alpha + b_n - c_n + 1)_{m_n} m_n!} \tag{27}$$

By substituting (27) into (20), we get (19).

**Lemma 1.** The multiple power series  $F_N^{(n)}$  is equal to the Lauricella D-hypergeometric function  $F_D^{(n)}$  (see Appendix A) [39] when  $a - c_n + 1 = c$  and is given as follows

$$F_N^{(n)}(a; b_1, \dots, b_n; c, c_n; x_1, \dots, x_n) = \sum_{\substack{m_1, \dots, \\ m_n=0}}^{+\infty} \frac{(a)_{\sum_{i=1}^n m_i} \prod_{i=1}^n (b_i)_{m_i}}{(a + b_n - c_n + 1)_{\sum_{i=1}^n m_i}} \prod_{i=1}^{n-1} \frac{x_i^{m_i}}{m_i!} \frac{(1 - x_n)^{m_n}}{m_n!} \tag{28}$$

$$= F_D^{(n)}(a, b_1, \dots, b_n; a + b_n - c_n + 1; x_1, \dots, x_{n-1}, 1 - x_n) \tag{29}$$

**Proof.** By using Equation (18) of the multiple power series  $F_N^{(n)}$  and after having simplified  $(a - c_n + 1)_{\sum_{i=1}^{n-1} m_i}$  to the numerator and  $(c)_{\sum_{i=1}^{n-1} m_i}$  to the denominator, we can get the result. □

### 3.3. Integral Representation for $F_N^{(n+1)}$

**Proposition 2.** The following integral representation is true for  $Real\{a\} > 0, Real\{a - c_{n+1} + 1\} > 0,$  and  $Real\{a - c_{n+1} + b_{n+1} + 1\} > 0$

$$\begin{aligned}
 & \frac{\Gamma(a)\Gamma(a - c_{n+1} + 1)}{\Gamma(a - c_{n+1} + b_{n+1} + 1)} F_N^{(n+1)}(a; b_1, \dots, b_{n+1}; c, c_{n+1}; x_1, \dots, x_{n+1}) \\
 &= \int_0^{\infty} e^{-r} r^{a-1} \Phi_2^{(n)}(b_1, \dots, b_n; c; rx_1, \dots, rx_n) U(b_{n+1}, c_{n+1}; rx_{n+1}) dr \tag{30}
 \end{aligned}$$

where  $U(\cdot)$  is the confluent hypergeometric function of the second kind (Section 9.21 in [35]) defined for  $Real\{b\} > 0, Real\{z\} > 0$  by the following integral representation

$$U(b, c; z) = \frac{1}{\Gamma(b)} \int_0^{\infty} e^{-zt} t^{b-1} (1 + t)^{c-b-1} dt \tag{31}$$

and  $\Phi_2^{(n)}(\cdot)$  is defined by Equation (6).

**Proof.** The multiple power series  $\Phi_2^{(n)}$  and the confluent hypergeometric function  $U(\cdot)$  are absolutely convergent on  $[0, +\infty]$ . Using these functions in the above integral and changing the order of integration and summation, which is easily justified by absolute convergence, we get

$$\begin{aligned}
 & \int_0^{\infty} e^{-r} r^{a-1} \Phi_2^{(n)}(b_1, \dots, b_n; c; rx_1, \dots, rx_n) U(b_{n+1}, c_{n+1}; rx_{n+1}) dr \\
 &= \sum_{m_1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \frac{(b_1)_{m_1} \dots (b_n)_{m_n}}{(c)_{\sum_{i=1}^n m_i}} \left( \prod_{i=1}^n \frac{x_i^{m_i}}{m_i!} \right) \mathbf{I} \tag{32}
 \end{aligned}$$

where integral  $\mathbf{I}$  is defined as follows

$$\mathbf{I} = \int_0^{\infty} e^{-r} r^{a-1 + \sum_{i=1}^n m_i} U(b_{n+1}, c_{n+1}; rx_{n+1}) dr. \tag{33}$$

Substituting the integral expression of  $U(\cdot)$  in the previous equation and replacing  $\alpha = a + \sum_{i=1}^n m_i$  to alleviate writing equations, we have

$$I = \frac{1}{\Gamma(b_{n+1})} \int_0^\infty \int_0^\infty \frac{e^{-(1+x_{n+1}t)r} r^{\alpha-1} t^{b_{n+1}-1}}{(1+t)^{-(c_{n+1}-b_{n+1}-1)}} dr dt. \tag{34}$$

Knowing that [35]

$$\int_0^\infty e^{-(1+x_{n+1}t)r} r^{\alpha-1} dr = \frac{\Gamma(\alpha)}{(1+x_{n+1}t)^\alpha} \tag{35}$$

and

$$\int_0^\infty \frac{t^{b_{n+1}-1} (1+t)^{c_{n+1}-b_{n+1}-1}}{(1+x_{n+1}t)^\alpha} dt = \frac{\Gamma(b_{n+1})\Gamma(\alpha - c_{n+1} + 1)}{\Gamma(\alpha + b_{n+1} - c_{n+1} + 1)} {}_2F_1(\alpha, b_{n+1}; \alpha + b_{n+1} - c_{n+1} + 1; 1 - x_{n+1}) \tag{36}$$

the new expression of  $I$  is then given by

$$I = \frac{\Gamma(\alpha)\Gamma(\alpha - c_{n+1} + 1)}{\Gamma(\alpha + b_{n+1} - c_{n+1} + 1)} \sum_{m_{n+1}=0}^{+\infty} \frac{(\alpha)_{m_{n+1}}(b_{n+1})_{m_{n+1}}}{(\alpha + b_{n+1} - c_{n+1} + 1)_{m_{n+1}}} \frac{(1 - x_{n+1})^{m_{n+1}}}{m_{n+1}!}. \tag{37}$$

Using the fact that  $\Gamma(\alpha) = \Gamma(a)(a)_{\sum_{i=1}^n m_i}$  and  $(a)_{\sum_{i=1}^n m_i} (a)_{m_{n+1}} = (a)_{\sum_{i=1}^{n+1} m_i}$ , and developing the same method to  $\Gamma(\alpha + b_{n+1} - c_{n+1} + 1)$ , the final complete expression of the integral is then given by

$$\begin{aligned} & \frac{\Gamma(a)\Gamma(a - c_{n+1} + 1)}{\Gamma(a + b_{n+1} - c_{n+1} + 1)} \sum_{m_1=0}^{\infty} \dots \sum_{m_{n+1}=0}^{\infty} \frac{(b_1)_{m_1} \dots (b_n)_{m_n}}{(c)_{\sum_{i=1}^n m_i}} \frac{(a - c_{n+1} + 1)_{\sum_{i=1}^n m_i} (b_{n+1})_{m_{n+1}} (a)_{\sum_{i=1}^{n+1} m_i}}{(a + b_{n+1} - c_{n+1} + 1)_{\sum_{i=1}^{n+1} m_i}} \prod_{i=1}^n \frac{x_i^{m_i}}{m_i!} \\ & \times \frac{(1 - x_{n+1})^{m_{n+1}}}{m_{n+1}!} = \frac{\Gamma(a)\Gamma(a - c_{n+1} + 1)}{\Gamma(a - c_{n+1} + b_{n+1} + 1)} F_N^{(n+1)}(a; b_1, \dots, b_{n+1}; c, c_{n+1}; x_1, \dots, x_{n+1}). \end{aligned} \tag{38}$$

#### 4. Expression of $E_{X^1} \{ \ln[1 + X^T \Sigma_1^{-1} X] \}$

**Proposition 3.** Let  $X^1$  be a random vector that follows a central MCD with pdf given by  $f_{X^1}(x|\Sigma_1, p)$ . Expectation  $E_{X^1} \{ \ln[1 + X^T \Sigma_1^{-1} X] \}$  is given as follows

$$E_{X^1} \{ \ln[1 + X^T \Sigma_1^{-1} X] \} = \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) \tag{39}$$

where  $\psi(\cdot)$  is the digamma function defined as the logarithmic derivative of the Gamma function (Section 8.36 in [35]).

**Proof.** Expectation  $E_{X^1} \{ \ln[1 + X^T \Sigma_1^{-1} X] \}$  is developed as follows

$$E_{X^1} \{ \ln[1 + X^T \Sigma_1^{-1} X] \} = \frac{A}{|\Sigma_1|^{\frac{1}{2}}} \int_{\mathbb{R}^p} \frac{\ln[1 + x^T \Sigma_1^{-1} x]}{[1 + x^T \Sigma_1^{-1} x]^{\frac{1+p}{2}}} dx \tag{40}$$

where  $A = \Gamma(\frac{1+p}{2})\pi^{-\frac{1+p}{2}}$ . Utilizing the following property  $\int \log(x)f(x)dx = \frac{\partial}{\partial a} \int x^a f(x) dx \Big|_{a=0}$ , as a consequence the expectation is given as follows

$$E_{X^1} \{ \ln[1 + X^T \Sigma_1^{-1} X] \} = \frac{A}{|\Sigma_1|^{\frac{1}{2}}} \frac{\partial}{\partial a} \int_{\mathbb{R}^p} [1 + x^T \Sigma_1^{-1} x]^{a-\frac{1+p}{2}} dx \Big|_{a=0} \tag{41}$$

Consider the transformation  $y = \Sigma_1^{-1/2}x$  where  $y = [y_1, y_2, \dots, y_p]^T$ . The Jacobian determinant is given by  $dy = |\Sigma_1|^{-1/2}dx$  (Theorem 1.12 in [40]). The new expression of the expectation is given by



$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{X}] \} = A \frac{\partial}{\partial a} \int_{\mathbb{R}^p} [1 + \mathbf{y}^T \mathbf{y}]^{a - \frac{1+p}{2}} d\mathbf{y} \Big|_{a=0}. \tag{42}$$

Let  $u = \mathbf{y}^T \mathbf{y}$  be a transformation where the Jacobian determinant is given by (Lemma 13.3.1 in [41])

$$d\mathbf{y} = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} u^{\frac{p}{2}-1} du. \tag{43}$$

The new expectation is as follows

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{X}] \} = \frac{\Gamma(\frac{1+p}{2})}{\pi^{1/2} \Gamma(\frac{p}{2})} \frac{\partial}{\partial a} \int_0^{+\infty} u^{\frac{p}{2}-1} (1+u)^{a - \frac{1+p}{2}} du \Big|_{a=0} \tag{44}$$

Using the definition of beta function, we can write that

$$\int_0^{+\infty} u^{\frac{p}{2}-1} (1+u)^{a - \frac{1+p}{2}} du = \frac{\Gamma(\frac{p}{2}) \Gamma(\frac{1}{2} - a)}{\Gamma(\frac{1+p}{2} - a)}. \tag{45}$$

The derivative of the last integral w.r.t  $a$  is given by

$$\frac{\partial}{\partial a} \int_0^{+\infty} u^{\frac{p}{2}-1} (1+u)^{a - \frac{1+p}{2}} du \Big|_{a=0} = \frac{\Gamma(\frac{p}{2}) \Gamma(\frac{1}{2})}{\Gamma(\frac{1+p}{2})} \left[ \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) \right] \tag{46}$$

Finally, the expression of  $E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{X}] \}$  is given by

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{X}] \} = \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right). \tag{47}$$

□

### 5. Expression of $E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \}$

**Proposition 4.** Let  $\mathbf{X}^1$  and  $\mathbf{X}^2$  be two random vectors that follow central MCDs with pdfs given, respectively, by  $f_{\mathbf{X}^1}(\mathbf{x} | \boldsymbol{\Sigma}_1, p)$  and  $f_{\mathbf{X}^2}(\mathbf{x} | \boldsymbol{\Sigma}_2, p)$ . Expectation  $E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \}$  is given as follows

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \} = \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) + \ln \lambda_p - \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p} \right) \right\} \Big|_{a=0}. \tag{48}$$

where  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of the real matrix  $\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}$ , and  $F_D^{(p)}(\cdot)$  represents the Lauricella  $D$ -hypergeometric function defined for  $p$  variables.

**Proof.** To prove Proposition 4, different steps are necessary. They are described in the following:

#### 5.1. First Step: Eigenvalue Expression

Expectation  $E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \}$  is computed as follows

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \} = \frac{A}{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}} \int_{\mathbb{R}^p} \frac{\ln[1 + \mathbf{x}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{x}]}{[1 + \mathbf{x}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{x}]^{\frac{1+p}{2}}} d\mathbf{x} \tag{49}$$



where  $A = \Gamma(\frac{1+p}{2})\pi^{-\frac{1+p}{2}}$ . Consider transformation  $\mathbf{y} = \Sigma_1^{-1/2}\mathbf{x}$  where  $\mathbf{y} = [y_1, y_2, \dots, y_p]^T$ . The Jacobian determinant is given by  $d\mathbf{y} = |\Sigma_1|^{-1/2}d\mathbf{x}$  (Theorem 1.12 in [40]) and matrix  $\Sigma = \Sigma_1^{\frac{1}{2}}\Sigma_2^{-1}\Sigma_1^{\frac{1}{2}}$  is a real symmetric matrix since  $\Sigma_1$  and  $\Sigma_2$  are real symmetric matrixes. Then, the expectation is evaluated as follows

$$E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\} = A \int_{\mathbb{R}^p} \frac{\ln[1 + \mathbf{y}^T \Sigma \mathbf{y}]}{[1 + \mathbf{y}^T \mathbf{y}]^{\frac{1+p}{2}}} d\mathbf{y}. \tag{50}$$

Matrix  $\Sigma$  can be diagonalized by an orthogonal matrix  $\mathbf{P}$  with  $\mathbf{P}^{-1} = \mathbf{P}^T$  and  $\Sigma = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$  where  $\mathbf{D}$  is a diagonal matrix composed of the eigenvalues of  $\Sigma$ . Considering that  $\mathbf{y}^T \Sigma \mathbf{y} = \text{tr}(\Sigma \mathbf{y} \mathbf{y}^T) = \text{tr}(\mathbf{P}\mathbf{D}\mathbf{P}^T \mathbf{y} \mathbf{y}^T) = \text{tr}(\mathbf{D}\mathbf{P}^T \mathbf{y} \mathbf{y}^T \mathbf{P})$ , the expectation can be written as

$$E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\} = A \int_{\mathbb{R}^p} \frac{\ln[1 + \text{tr}(\mathbf{D}\mathbf{P}^T \mathbf{y} \mathbf{y}^T \mathbf{P})]}{[1 + \mathbf{y}^T \mathbf{y}]^{\frac{1+p}{2}}} d\mathbf{y}. \tag{51}$$

Let  $\mathbf{z} = \mathbf{P}^T \mathbf{y}$  with  $\mathbf{z} = [z_1, z_2, \dots, z_p]^T$  be a transformation where the Jacobian determinant is given by  $d\mathbf{z} = |\mathbf{P}^T|d\mathbf{y} = d\mathbf{y}$ . Using the fact that  $\text{tr}(\mathbf{D}\mathbf{P}^T \mathbf{y} \mathbf{y}^T \mathbf{P}) = \text{tr}(\mathbf{D}\mathbf{z}\mathbf{z}^T) = \mathbf{z}^T \mathbf{D}\mathbf{z}$  and  $\mathbf{y}^T \mathbf{y} = \mathbf{z}^T \mathbf{P}^T \mathbf{P}\mathbf{z} = \mathbf{z}^T \mathbf{z}$ , then the previous expectation (51) is given as follows

$$E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\} = A \int_{\mathbb{R}^p} \frac{\ln[1 + \mathbf{z}^T \mathbf{D}\mathbf{z}]}{[1 + \mathbf{z}^T \mathbf{z}]^{\frac{1+p}{2}}} d\mathbf{z} \tag{52}$$

$$= A \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \frac{\ln[1 + \sum_{i=1}^p \lambda_i z_i^2]}{[1 + \sum_{i=1}^p z_i^2]^{\frac{1+p}{2}}} dz_1 \dots dz_p \tag{53}$$

where  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of  $\Sigma_1 \Sigma_2^{-1}$ .

5.2. Second Step: Polar Decomposition

Let the independent real variables  $z_1, \dots, z_p$  be transformed to the general polar coordinates  $r, \theta_1, \dots, \theta_{p-1}$  as follows, where  $r > 0, -\pi/2 < \theta_j \leq \pi/2, j = 1, \dots, p - 2, -\pi < \theta_{p-1} \leq \pi$  [40],

$$z_1 = r \sin \theta_1 \tag{54}$$

$$z_2 = r \cos \theta_1 \sin \theta_2 \tag{55}$$

$$z_j = r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{j-1} \sin \theta_j, \quad j = 2, 3, \dots, p - 1 \tag{56}$$

$$z_p = r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{p-1}. \tag{57}$$

The Jacobian determinant according to theorem (1.24) in [40] is

$$dz_1 \dots dz_p = r^{p-1} \prod_{j=1}^{p-1} |\cos \theta_j|^{p-j-1} dr d\theta_j. \tag{58}$$

It is clear that with the last transformations, we get  $\sum_{i=1}^p z_i^2 = r^2$  and the multiple integral in (53) is then given as follows

$$E_{\mathbf{X}^1}\{\ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}]\} = A \int_0^{+\infty} \frac{r^{p-1}}{[1 + r^2]^{\frac{1+p}{2}}} \int_{-\pi/2}^{\pi/2} \dots \int_{-\pi}^{\pi} \left( \prod_{j=1}^{p-1} |\cos \theta_j|^{p-j-1} \right) \times \ln [1 + r^2(\lambda_1 \sin^2 \theta_1 + \dots + \lambda_p \cos^2 \theta_1 \dots \cos^2 \theta_{p-1})] dr \prod_{j=1}^{p-1} d\theta_j. \tag{59}$$

By replacing the expression of  $\sin^2 \theta_j$  by  $1 - \cos^2 \theta_j$ , for  $j = 1, \dots, p - 1$ , we have the following expression

$$\begin{aligned} \lambda_1 \sin^2 \theta_1 + \dots + \lambda_p \cos^2 \theta_1 \dots \cos^2 \theta_{p-1} &= \lambda_1 + (\lambda_2 - \lambda_1) \cos^2 \theta_1 \\ &+ \dots + (\lambda_p - \lambda_{p-1}) \cos^2 \theta_1 \cos^2 \theta_2 \dots \cos^2 \theta_{p-1}. \end{aligned} \tag{60}$$

Let  $x_i = \cos^2 \theta_i$  be a transformation to use where  $dx_i = 2x_i^{1/2}(1 - x_i)^{1/2}d\theta_i$ . Then the expectation given by the multiple integral over all  $\theta_j, j = 1, \dots, p - 1$  is as follows

$$2A \int_0^{+\infty} \frac{r^{p-1}}{[1+r^2]^{\frac{1+p}{2}}} \int_0^1 \dots \int_0^1 \left( \prod_{j=1}^{p-1} x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} \right) \ln[1+r^2 B_p(x_1, \dots, x_{p-1})] dr dx_1 \dots dx_{p-1} \tag{61}$$

where  $B_p(x_1, \dots, x_{p-1}) = \lambda_1 + (\lambda_2 - \lambda_1)x_1 + \dots + (\lambda_p - \lambda_{p-1})x_1x_2 \dots x_{p-1}$ ,  $p \geq 1$  and  $B_1 = \lambda_1$ . In the following, we use the notation  $B_p$  instead of  $B_p(x_1, \dots, x_{p-1})$  to alleviate writing equations.

Let  $t = r^2$  be transformation to use. Then, one can write

$$= A \int_0^{+\infty} \frac{t^{\frac{p}{2}-1}}{[1+t]^{\frac{1+p}{2}}} \int_0^1 \dots \int_0^1 \left( \prod_{j=1}^{p-1} x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} \right) \ln[1+tB_p] dt dx_1 \dots dx_{p-1}. \tag{62}$$

In order to solve the integral in (62), we consider the following property given by  $\int \log(x)f(x) dx = -\frac{\partial}{\partial a} \int x^{-a}f(x)dx|_{a=0}$  and the following equation given as follows

$$(1+B_p t)^{-a} = \frac{1}{\Gamma(a)} \int_0^{+\infty} y^{a-1} e^{-(1+B_p t)y} dy. \tag{63}$$

Making use of the above equation, we obtain a new expression of (62) given as follows

$$\begin{aligned} E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}] \} \\ = -\frac{\partial}{\partial a} \left\{ \frac{A}{\Gamma(a)} \int_0^{+\infty} \frac{t^{\frac{p}{2}-1}}{[1+t]^{\frac{1+p}{2}}} \int_0^{+\infty} y^{a-1} e^{-(1+B_p t)y} \int_0^1 \dots \int_0^1 \prod_{j=1}^{p-1} x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} dx_j dy dt \right\} \Big|_{a=0} \end{aligned} \tag{64}$$

$$= -\frac{\partial}{\partial a} \left\{ \frac{A}{\Gamma(a)} \int_0^{+\infty} \frac{t^{\frac{p}{2}-1}}{[1+t]^{\frac{1+p}{2}}} \int_0^{+\infty} y^{a-1} e^{-y} \mathbf{H}(t, y) dy dt \right\} \Big|_{a=0} \tag{65}$$

where  $\mathbf{H}(t, y)$  is defined as

$$\mathbf{H}(t, y) = \int_0^1 \dots \int_0^1 e^{-B_p t y} \prod_{j=1}^{p-1} x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} dx_j. \tag{66}$$

### 5.3. Third Step: Expression for $H(t,y)$ by Humbert and Beta Functions

Let  $x'_i = 1 - x_i, i = 1, \dots, p - 1$  be transformations to use. Then

$$(\lambda_2 - \lambda_1)x_1 = (\lambda_2 - \lambda_1)(1 - x'_1) \tag{67}$$

$$(\lambda_3 - \lambda_2)x_1x_2 = (\lambda_3 - \lambda_2)(1 - x'_1)[1 - x'_2] \tag{68}$$

$$(\lambda_4 - \lambda_3)x_1x_2x_3 = (\lambda_4 - \lambda_3)(1 - x'_1)(1 - x'_2)[1 - x'_3] \tag{69}$$

$\vdots = \vdots$

$$(\lambda_p - \lambda_{p-1}) \prod_{i=1}^{p-1} x_i = (\lambda_p - \lambda_{p-1}) \prod_{i=1}^{p-1} (1 - x'_i). \tag{70}$$

Adding equations from (67) to (70), we can state that the new expression of the function  $B_p$  becomes

$$B_p = \lambda_p - (\lambda_p - \lambda_1)x'_1 - (\lambda_p - \lambda_2)(1 - x'_1)x'_2 - (\lambda_p - \lambda_3)(1 - x'_1)(1 - x'_2)x'_3 - \dots - (\lambda_p - \lambda_{p-1})(1 - x'_1) \dots (1 - x'_{p-2})x'_{p-1}. \tag{71}$$

Then, the multiple integral  $\mathbf{H}(t, y)$  given by (66) can be written otherwise

$$\mathbf{H}(t, y) = \int_0^1 \dots \int_0^1 e^{-B_p t y} \prod_{j=1}^{p-1} (1 - x'_j)^{\frac{p-j}{2}-1} x_j'^{-\frac{1}{2}} dx'_1 \dots dx'_{p-1}. \tag{72}$$

Let the real variables  $x'_1, x'_2, \dots, x'_{p-1}$  be transformed to the real variables  $u_1, u_2, \dots, u_{p-1}$  as follows

$$u_1 = x'_1 \tag{73}$$

$$u_2 = (1 - x'_1)x'_2 = (1 - u_1)x'_2 \tag{74}$$

$$u_3 = (1 - x'_1)(1 - x'_2)x'_3 = (1 - u_1 - u_2)x'_3 \tag{75}$$

⋮

$$u_{p-1} = \prod_{i=1}^{p-2} (1 - x'_i)x'_{p-1} = (1 - \sum_{i=1}^{p-2} u_i)x'_{p-1}. \tag{76}$$

The Jacobian determinant is given by

$$du_1 \dots du_{p-1} = \prod_{j=1}^{p-1} \left(1 - \sum_{i=1}^{j-1} u_i\right) dx'_1 \dots dx'_{p-1}. \tag{77}$$

Accordingly, the new expression of  $B_p$  becomes

$$B_p = \lambda_p - \sum_{i=1}^{p-1} (\lambda_p - \lambda_i)u_i. \tag{78}$$

As a consequence, the new domain of the multiple integral (72) is  $\Delta = \{(u_1, u_2, \dots, u_{p-1}) \in \mathbb{R}^{p-1}; 0 \leq u_1 \leq 1, 0 \leq u_2 \leq 1 - u_1, 0 \leq u_3 \leq 1 - u_1 - u_2, \dots, \text{and } 0 \leq u_{p-1} \leq 1 - u_1 - u_2 \dots - u_{p-2}\}$ , and the expression of  $\mathbf{H}(t, y)$  is given as follows

$$\mathbf{H}(t, y) = \int \dots \int_{\Delta} e^{-B_p t y} \prod_{j=1}^{p-1} \left(1 - \sum_{i=1}^{j-1} u_i\right)^{-1} \left(\frac{u_j}{1 - \sum_{i=1}^{j-1} u_i}\right)^{-\frac{1}{2}} \prod_{j=1}^{p-1} \left(1 - \frac{u_j}{1 - \sum_{i=1}^{j-1} u_i}\right)^{\frac{p-j}{2}-1} du_j \tag{79}$$

$$= \int \dots \int_{\Delta} e^{-B_p t y} \prod_{j=1}^{p-1} u_j^{-\frac{1}{2}} \left(1 - \sum_{i=1}^j u_i\right)^{\frac{p-j}{2}-1} \left(1 - \sum_{i=1}^{j-1} u_i\right)^{\frac{1}{2}-\frac{p-j}{2}} du_1 \dots du_{p-1} \tag{80}$$

$$= \int \dots \int_{\Delta} e^{-B_p t y} \left(1 - \sum_{i=1}^{p-1} u_i\right)^{\frac{p}{2}-\frac{p-1}{2}-1} \prod_{j=1}^{p-1} u_j^{-\frac{1}{2}} du_j \tag{81}$$

$$= e^{-\lambda_p t y} \int \dots \int_{\Delta} \left(1 - \sum_{i=1}^{p-1} u_i\right)^{-\frac{1}{2}} \prod_{i=1}^{p-1} u_i^{-\frac{1}{2}} e^{(\lambda_p - \lambda_i)u_i t y} du_i. \tag{82}$$

Using Proposition 1, we subsequently find that

$$\mathbf{H}(t, y) = e^{-\lambda_p t y} B \left(\underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p\right) \Phi_2^{(p-1)} \left(\underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{p-1}; \frac{p}{2}; (\lambda_p - \lambda_1)ty, (\lambda_p - \lambda_2)ty, \dots, (\lambda_p - \lambda_{p-1})ty\right). \tag{83}$$

where  $\Phi_2^{(p-1)}(\cdot)$  is the Humbert series of  $p - 1$  variables and  $B(\frac{1}{2}, \dots, \frac{1}{2})$  is the multivariate beta function. Applying the following successive two transformations  $r = ty$  ( $dr = tdy$ ) and  $u = 1/t$  ( $du = -u^2dt$ ), the new expression of the expectation given by (65) is written as follows

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}] \} = -\frac{\partial}{\partial a} \left\{ \frac{A}{\Gamma(a)} B\left(\underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p\right) \int_0^{+\infty} r^{a-1} e^{-\lambda_p r} \right. \\ \left. \times \Phi_2^{(p-1)}\left(\underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_{p-1}; \frac{1}{2}; (\lambda_p - \lambda_1)r, \dots, (\lambda_p - \lambda_{p-1})r\right) \left( \int_0^{+\infty} u^{a-\frac{1}{2}} (1+u)^{-\frac{1+p}{2}} e^{-ru} du \right) dr \right\} \Big|_{a=0}. \tag{84}$$

5.4. Final Step

The last integral is related to the confluent hypergeometric function of the second kind  $U(\cdot)$  as follows

$$\int_0^{+\infty} u^{a-\frac{1}{2}} (1+u)^{-\frac{1+p}{2}} e^{-ru} du = \Gamma\left(a + \frac{1}{2}\right) U\left(a + \frac{1}{2}, a + 1 - \frac{p}{2}, r\right). \tag{85}$$

As a consequence, the new expression is

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}] \} = -\frac{\partial}{\partial a} \left\{ A \frac{\Gamma\left(a + \frac{1}{2}\right)}{\Gamma(a)} B\left(\frac{1}{2}, \dots, \frac{1}{2}\right) \right. \\ \left. \times \int_0^{+\infty} r^{a-1} e^{-\lambda_p r} \Phi_2^{(p-1)}\left(\frac{1}{2}, \dots, \frac{1}{2}; \frac{1}{2}; 1; (\lambda_p - \lambda_1)r, \dots, (\lambda_p - \lambda_{p-1})r\right) U\left(a + \frac{1}{2}, a + 1 - \frac{p}{2}, r\right) dr \right\} \Big|_{a=0}. \tag{86}$$

Using the transformation  $r' = \lambda_p r$  and the Proposition 2, and taking into account the expression of  $A$ , the new expression becomes

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}] \} = -\frac{\partial}{\partial a} \left\{ \frac{B\left(a + \frac{1}{2}, \frac{p}{2}\right)}{B\left(\frac{1}{2}, \frac{p}{2}\right)} \lambda_p^{-a} \right. \\ \left. \times F_N^{(p)}\left(a; \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; \frac{p}{2}, a - \frac{p}{2} + 1; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, \lambda_p^{-1}\right) \right\} \Big|_{a=0} \tag{87}$$

Knowing that

$$\frac{\partial}{\partial a} \left\{ \frac{B\left(\frac{p}{2}, a + \frac{1}{2}\right)}{B\left(\frac{p}{2}, \frac{1}{2}\right)} \right\} \Big|_{a=0} = \psi\left(\frac{1}{2}\right) - \psi\left(\frac{1+p}{2}\right), \text{ and} \tag{88}$$

$$F_N^{(p)}\left(a; \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; \frac{p}{2}, a - \frac{p}{2} + 1; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, \lambda_p^{-1}\right) \Big|_{a=0} = 1, \tag{89}$$

the new expression of  $E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}] \}$  becomes

$$E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \Sigma_2^{-1} \mathbf{X}] \} = \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) \\ - \frac{\partial}{\partial a} \left\{ \lambda_p^{-a} F_N^{(p)}\left(a; \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; \frac{p}{2}, a - \frac{p}{2} + 1; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, \lambda_p^{-1}\right) \right\} \Big|_{a=0}. \tag{90}$$

Applying the expression given by (18) of Definition 2 and relying on Lemma 1, the final result corresponds to the D-hypergeometric function of Lauricella  $F_D^{(p)}(\cdot)$  given by

$$\begin{aligned}
 E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \} &= \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) \\
 &- \frac{\partial}{\partial a} \left\{ \lambda_p^{-a} \sum_{\substack{m_1, \dots, \\ m_p=0}}^{+\infty} \frac{(a)_{\sum_{i=1}^p m_i} (a + \frac{1}{2})_{m_p} \prod_{i=1}^{p-1} (\frac{1}{2})_{m_i}}{(a + \frac{1+p}{2})_{\sum_{i=1}^p m_i}} \prod_{i=1}^{p-1} \left(1 - \frac{\lambda_i}{\lambda_p}\right)^{m_i} \frac{1}{m_i!} \frac{(1 - \lambda_p^{-1})^{m_p}}{m_p!} \right\} \Big|_{a=0} \quad (91)
 \end{aligned}$$

$$= \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) - \frac{\partial}{\partial a} \left\{ \lambda_p^{-a} F_D^{(p)}\left(a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p}\right) \right\} \Big|_{a=0}. \quad (92)$$

The final development of the previous expression is as follows

$$\begin{aligned}
 E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \} &= \psi\left(\frac{1+p}{2}\right) - \psi\left(\frac{1}{2}\right) + \ln \lambda_p \\
 &- \frac{\partial}{\partial a} \left\{ F_D^{(p)}\left(a, \underbrace{\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p}\right) \right\} \Big|_{a=0}. \quad (93)
 \end{aligned}$$

□

In this section, we presented the exact expression of  $E_{\mathbf{X}^1} \{ \ln[1 + \mathbf{X}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{X}] \}$ . In addition, the multiple power series  $F_D^{(p)}$  which appears to be a special case of  $F_N^{(p)}$  provides many properties and numerous transformations (see Appendix A) that make easier the convergence of the multiple power series. In the next section, we establish the KLD closed-form expression based on the expression of the latter expectation.

### 6. KLD between Two Central MCDs

Plugging (39) and (93) into (5) yields the closed-form expression of the KLD between two central MCDs with pdfs  $f_{\mathbf{X}^1}(\mathbf{x}|\boldsymbol{\Sigma}_1, p)$  and  $f_{\mathbf{X}^2}(\mathbf{x}|\boldsymbol{\Sigma}_2, p)$ . This result is presented in the following theorem.

**Theorem 1.** Let  $\mathbf{X}^1$  and  $\mathbf{X}^2$  be two random vectors that follow central MCDs with pdfs given, respectively, by  $f_{\mathbf{X}^1}(\mathbf{x}|\boldsymbol{\Sigma}_1, p)$  and  $f_{\mathbf{X}^2}(\mathbf{x}|\boldsymbol{\Sigma}_2, p)$ . The Kullback–Leibler divergence between central MCDs is

$$\begin{aligned}
 \text{KL}(\mathbf{X}^1 || \mathbf{X}^2) &= -\frac{1}{2} \log \prod_{i=1}^p \lambda_i + \frac{1+p}{2} \left[ \log \lambda_p \right. \\
 &- \left. \frac{\partial}{\partial a} \left\{ F_D^{(p)}\left(a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p}\right) \right\} \Big|_{a=0} \right] \quad (94)
 \end{aligned}$$

where  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of the real matrix  $\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}$ , and  $F_D^{(p)}(\cdot)$  represents the Lauricella D-hypergeometric function defined for  $p$  variables.

Lauricella [39] gave several transformation formulas (see Appendix A), whose relations (A5)–(A7), and (A9) are applied to  $F_D^{(p)}(\cdot)$  in (94). The results of transformation are as follows

$$\begin{aligned}
 &F_D^{(p)}\left(a, \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p}\right) \\
 &= \lambda_p^{a+\frac{p}{2}} \prod_{i=1}^{p-1} \lambda_i^{-\frac{1}{2}} F_D^{(p)}\left(\frac{1+p}{2}, \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_p}{\lambda_1}, \dots, 1 - \frac{\lambda_p}{\lambda_{p-1}}, 1 - \lambda_p\right) \tag{95}
 \end{aligned}$$

$$= \left(\frac{\lambda_1}{\lambda_p}\right)^{-a} F_D^{(p)}\left(a, \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_p}{\lambda_1}, \dots, 1 - \frac{\lambda_2}{\lambda_1}, 1 - \frac{1}{\lambda_1}\right) \tag{96}$$

$$= \lambda_p^a F_D^{(p)}\left(a, \frac{1}{2}, \dots, \frac{1}{2}; a + \frac{1+p}{2}; 1 - \lambda_1, 1 - \lambda_2, \dots, 1 - \lambda_p\right) \tag{97}$$

$$= \lambda_p^a \prod_{i=1}^p \lambda_i^{-\frac{1}{2}} F_D^{(p)}\left(\frac{1+p}{2}, \frac{1}{2}, \dots, \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, 1 - \frac{1}{\lambda_2}, \dots, 1 - \frac{1}{\lambda_p}\right). \tag{98}$$

Considering the above equations, it is easy to provide different expressions of  $KL(\mathbf{X}^1 || \mathbf{X}^2)$  shown in Table 1. The derivative of the Lauricella D-hypergeometric series with respect to  $a$  goes through the derivation of the following expression

$$\left. \frac{\partial}{\partial a} \left\{ F_D^{(p)}\left(a, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p}\right) \right\} \right|_{a=0} \tag{99}$$

$$= \sum_{\substack{m_1, \dots, \\ m_p=0}}^{+\infty} \frac{\partial}{\partial a} \left\{ \frac{(a)_{\sum_{i=1}^p m_i} (a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_{\sum_{i=1}^p m_i}} \right\} \bigg|_{a=0} \prod_{i=1}^{p-1} \binom{1}{2}_{m_i} \left(1 - \frac{\lambda_i}{\lambda_p}\right)^{m_i} \frac{1}{m_i!} \frac{(1 - \lambda_p^{-1})^{m_p}}{m_p!} \tag{100}$$

The derivative with respect to  $a$  of the Lauricella D-hypergeometric series and its transformations goes through the following expressions (see Appendix B for demonstration)

$$\left. \frac{\partial}{\partial a} \left\{ \frac{(a)_{\sum_{i=1}^p m_i} (a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_{\sum_{i=1}^p m_i}} \right\} \right|_{a=0} = \frac{(\frac{1}{2})_{m_p} (1)_{\sum_{i=1}^p m_i}}{(\frac{1+p}{2})_{\sum_{i=1}^p m_i} (\sum_{i=1}^p m_i)}, \tag{101}$$

$$\left. \frac{\partial}{\partial a} \left\{ \frac{(a)_{\sum_{i=1}^p m_i}}{(a + \frac{1+p}{2})_{\sum_{i=1}^p m_i}} \right\} \right|_{a=0} = \frac{(1)_{\sum_{i=1}^p m_i}}{(\frac{1+p}{2})_{\sum_{i=1}^p m_i} (\sum_{i=1}^p m_i)}, \tag{102}$$

$$\left. \frac{\partial}{\partial a} \left\{ \frac{(a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_{\sum_{i=1}^p m_i}} \right\} \right|_{a=0} = \frac{(\frac{1}{2})_{m_p}}{(\frac{1+p}{2})_{\sum_{i=1}^p m_i}} \left[ \sum_{k=0}^{m_p-1} \frac{1}{k + \frac{1}{2}} - \sum_{k=0}^{\sum_{i=1}^p m_i-1} \frac{1}{k + \frac{1+p}{2}} \right], \tag{103}$$

$$\left. \frac{\partial}{\partial a} \left\{ \frac{1}{(a + \frac{1+p}{2})_{\sum_{i=1}^p m_i}} \right\} \right|_{a=0} = \frac{-1}{(\frac{1+p}{2})_{\sum_{i=1}^p m_i}} \sum_{k=0}^{\sum_{i=1}^p m_i-1} \frac{1}{k + \frac{1+p}{2}}. \tag{104}$$

To derive the closed-form expression of  $d_{KL}(\mathbf{X}^1, \mathbf{X}^2)$  we have to evaluate the expression of  $KL(\mathbf{X}^2 || \mathbf{X}^1)$ . The latter can be easily deduced from  $KL(\mathbf{X}^1 || \mathbf{X}^2)$  as follows

$$\begin{aligned}
 KL(\mathbf{X}^2 || \mathbf{X}^1) &= \frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \left[ \log \lambda_p \right. \\
 &\left. + \frac{\partial}{\partial a} \left\{ F_D^{(p)}\left(a, \frac{1}{2}, \dots, \frac{1}{2}, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_p}{\lambda_1}, \dots, 1 - \frac{\lambda_p}{\lambda_{p-1}}, 1 - \lambda_p\right) \right\} \bigg|_{a=0} \right]. \tag{105}
 \end{aligned}$$

Proceeding in the same way by using Lauricella transformations, different expressions of  $KL(\mathbf{X}^2 || \mathbf{X}^1)$  are provided in Table 1. Finally, given the above results, it is straightforward to compute the symmetric KL similarity measure  $d_{KL}(\mathbf{X}^1, \mathbf{X}^2)$  between  $\mathbf{X}^1$  and  $\mathbf{X}^2$ . Technically, any combination of the  $KL(\mathbf{X}^1 || \mathbf{X}^2)$  and  $KL(\mathbf{X}^2 || \mathbf{X}^1)$  expressions is possible to compute  $d_{KL}(\mathbf{X}^1, \mathbf{X}^2)$ . However, we choose the same convergence region for the two divergences for the calculation of the distance. Some expressions of  $d_{KL}(\mathbf{X}^1, \mathbf{X}^2)$  are given in Table 1.

**Table 1.** KLD and KL distance computed when  $\mathbf{X}^1$  and  $\mathbf{X}^2$  are two random vectors following central MCDs with pdfs  $f_{\mathbf{X}^1}(\mathbf{x}|\Sigma_1, p)$  and  $f_{\mathbf{X}^2}(\mathbf{x}|\Sigma_2, p)$ .

$\begin{aligned} & \text{KL}(\mathbf{X}^1  \mathbf{X}^2) \\ &= -\frac{1}{2} \log \prod_{i=1}^p \lambda_i + \frac{1+p}{2} \left[ \log \lambda_p - \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} \right] \end{aligned} \tag{106}$	
$= -\frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \lambda_p^{\frac{p}{2}} \prod_{i=1}^{p-1} \lambda_i^{-\frac{1}{2}} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_p}{\lambda_1}, \dots, 1 - \frac{\lambda_p}{\lambda_{p-1}}, 1 - \lambda_p \right) \right\} \Big _{a=0} \tag{107}$	
$= -\frac{1}{2} \log \prod_{i=1}^p \lambda_i + \frac{1+p}{2} \left[ \log \lambda_1 - \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_p}{\lambda_1}, \dots, 1 - \frac{\lambda_2}{\lambda_1}, 1 - \frac{1}{\lambda_1} \right) \right\} \Big _{a=0} \right] \tag{108}$	
$= -\frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \lambda_1, \dots, 1 - \lambda_p \right) \right\} \Big _{a=0} \tag{109}$	
$= -\frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \prod_{i=1}^p \lambda_i^{-\frac{1}{2}} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, \dots, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} \tag{110}$	
$\begin{aligned} & \text{KL}(\mathbf{X}^2  \mathbf{X}^1) \\ &= \frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \left[ \log \lambda_p + \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_p}{\lambda_1}, \dots, 1 - \frac{\lambda_p}{\lambda_{p-1}}, 1 - \lambda_p \right) \right\} \Big _{a=0} \right] \end{aligned} \tag{111}$	
$= \frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \lambda_p^{\frac{p}{2}} \prod_{i=1}^{p-1} \lambda_i^{\frac{1}{2}} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} \tag{112}$	
$= \frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \left[ \log \lambda_1 + \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_1}{\lambda_2}, 1 - \lambda_1 \right) \right\} \Big _{a=0} \right] \tag{113}$	
$= \frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, \dots, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} \tag{114}$	
$= \frac{1}{2} \log \prod_{i=1}^p \lambda_i - \frac{1+p}{2} \prod_{i=1}^p \lambda_i^{\frac{1}{2}} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \lambda_1, \dots, 1 - \lambda_p \right) \right\} \Big _{a=0} \tag{115}$	
$\begin{aligned} & d_{\text{KL}}(\mathbf{X}^1, \mathbf{X}^2) \\ &= \frac{1+p}{2} \left[ \log \lambda_p - \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} - \lambda_p^{-\frac{p}{2}} \prod_{i=1}^{p-1} \lambda_i^{\frac{1}{2}} \right. \\ & \quad \left. \times \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p, a + \frac{1}{2}; a + \frac{1+p}{2}; 1 - \frac{\lambda_1}{\lambda_p}, \dots, 1 - \frac{\lambda_{p-1}}{\lambda_p}, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} \right] \end{aligned} \tag{116}$	
$= -\frac{1+p}{2} \left[ \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \lambda_1, \dots, 1 - \lambda_p \right) \right\} \Big _{a=0} + \prod_{i=1}^p \lambda_i^{\frac{1}{2}} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \lambda_1, \dots, 1 - \lambda_p \right) \right\} \Big _{a=0} \right] \tag{117}$	
$= -\frac{1+p}{2} \left[ \prod_{i=1}^p \lambda_i^{-\frac{1}{2}} \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( \frac{1+p}{2}, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, \dots, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} + \frac{\partial}{\partial a} \left\{ F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \frac{1}{\lambda_1}, \dots, 1 - \frac{1}{\lambda_p} \right) \right\} \Big _{a=0} \right] \tag{118}$	

**7. Particular Cases: Univariate and Bivariate Cauchy Distribution**

**7.1. Case of  $p = 1$**

This case corresponds to the univariate Cauchy distribution. The KLD is given by



$$KL(X^1||X^2) = -\frac{1}{2} \log \lambda - \frac{\partial}{\partial a} \left\{ {}_2F_1\left(a, \frac{1}{2}; a + 1; 1 - \lambda\right) \right\} \Big|_{a=0} \tag{119}$$

where  ${}_2F_1$  is the Gauss’s hypergeometric function. The expression of the derivative of  ${}_2F_1$  is given as follows (see Appendix C.1 for details of computation)

$$\begin{aligned} \frac{\partial}{\partial a} \left\{ {}_2F_1\left(a, \frac{1}{2}; a + 1; 1 - \lambda\right) \right\} \Big|_{a=0} &= \sum_{n=1}^{\infty} \binom{1}{\frac{1}{2}}_n \frac{1(1-\lambda)^n}{n!} \\ &= -2 \ln \left( \frac{1 + \lambda^{1/2}}{2} \right). \end{aligned} \tag{120}$$

Accordingly, the KLD is then expressed as

$$KL(X^1||X^2) = \log \frac{(1 + \lambda^{\frac{1}{2}})^2}{4\lambda^{\frac{1}{2}}} \tag{121}$$

$$= \log \frac{(1 + \lambda^{-\frac{1}{2}})^2}{4\lambda^{-\frac{1}{2}}} = KL(X^2||X^1). \tag{122}$$

We conclude that KLD between Cauchy densities is always symmetric. Interestingly, this is consistent with the result presented in [31].

7.2. Case of  $p = 2$

This case corresponds to the Bivariate Cauchy distribution. The KLD is then given by

$$KL(\mathbf{X}^1||\mathbf{X}^2) = -\frac{1}{2} \log \lambda_1 \lambda_2 - \frac{3}{2} \frac{\partial}{\partial a} \left\{ F_1\left(a, \frac{1}{2}, \frac{1}{2}; a + \frac{3}{2}; 1 - \lambda_1, 1 - \lambda_2\right) \right\} \Big|_{a=0} \tag{123}$$

where  $F_1$  is the Appell’s hypergeometric function (see Appendix A). The expression of the derivative of  $F_1$  can be further developed

$$\begin{aligned} &\frac{\partial}{\partial a} \left\{ F_1\left(a, \frac{1}{2}, \frac{1}{2}; a + \frac{3}{2}; 1 - \lambda_1, 1 - \lambda_2\right) \right\} \Big|_{a=0} \\ &= \sum_{n,m=0}^{+\infty} \frac{(1)_{m+n} (\frac{1}{2})_n (\frac{1}{2})_m}{(\frac{3}{2})_{m+n}} \frac{1}{m+n} \frac{(1-\lambda_1)^n}{n!} \frac{(1-\lambda_2)^m}{m!}. \end{aligned} \tag{124}$$

In addition, when the eigenvalue  $\lambda_i$  for  $i = 1, 2$  takes some particular values, the expression of the KLD becomes very simple. In the following, we show some cases:

$(\lambda_1 = 1, \lambda_2 \neq 1)$  or  $(\lambda_2 = 1, \lambda_1 \neq 1)$

For this particular case, we have

$$\frac{\partial}{\partial a} \left\{ F_1\left(a, \frac{1}{2}, \frac{1}{2}; a + \frac{3}{2}; 1 - \lambda_i, 0\right) \right\} \Big|_{a=0} = \frac{\partial}{\partial a} \left\{ {}_2F_1\left(a, \frac{1}{2}; a + \frac{3}{2}; 1 - \lambda_i\right) \right\} \Big|_{a=0} \tag{125}$$

$$= -\ln \lambda_i + \frac{1}{\sqrt{1-\lambda_i}} \ln \left( \frac{1 - \sqrt{1-\lambda_i}}{1 + \sqrt{1-\lambda_i}} \right) + 2. \tag{126}$$

The demonstration of the derivation is shown in Appendix C.2. Then, KLD becomes equal to

$$KL(X^1||X^2) = \ln \lambda_i - \frac{3}{2} \frac{1}{\sqrt{1-\lambda_i}} \ln \left( \frac{1 - \sqrt{1-\lambda_i}}{1 + \sqrt{1-\lambda_i}} \right) - 3. \tag{127}$$

$$\lambda_1 = \lambda_2 = \lambda$$

For this particular case, we have

$$\left. \frac{\partial}{\partial a} \left\{ F_1 \left( a, \frac{1}{2}, \frac{1}{2}; a + \frac{3}{2}; 1 - \lambda, 1 - \lambda \right) \right\} \right|_{a=0} = \left. \frac{\partial}{\partial a} \left\{ {}_2F_1 \left( a, 1; a + \frac{3}{2}; 1 - \lambda \right) \right\} \right|_{a=0} \tag{128}$$

$$= -\frac{2}{\sqrt{1-\lambda^{-1}}} \ln(\sqrt{\lambda} + \sqrt{\lambda-1}) + 2. \tag{129}$$

For more details about the demonstration see Appendix C.3. The KLD becomes equal to

$$\text{KL}(\mathbf{X}^1 || \mathbf{X}^2) = -\ln \lambda + \frac{3}{\sqrt{1-\lambda^{-1}}} \ln(\sqrt{\lambda} + \sqrt{\lambda-1}) - 3. \tag{130}$$

It is easy to deduce that

$$\text{KL}(\mathbf{X}^2 || \mathbf{X}^1) = \ln \lambda + \frac{3}{\sqrt{1-\lambda}} \ln(\sqrt{\lambda^{-1}} + \sqrt{\lambda^{-1}-1}) - 3. \tag{131}$$

This result can be demonstrated using the same process as  $\text{KL}(\mathbf{X}^1 || \mathbf{X}^2)$ . It is worth to notice that  $\text{KL}(\mathbf{X}^1 || \mathbf{X}^2) \neq \text{KL}(\mathbf{X}^2 || \mathbf{X}^1)$  which leads us to conclude that the property of symmetry observed for the univariate case is no longer valid in the multivariate case. Nielsen et al. in [32] gave the same conclusion.

### 8. Implementation and Comparison with Monte Carlo Technique

In this section, we show how we practically compute the numerical values of the KLD, especially when we have several equivalent expressions which differ in the region of convergence. To reach this goal, the eigenvalues of  $\Sigma_1 \Sigma_2^{-1}$  are rearranged in a descending order  $\lambda_p > \lambda_{p-1} > \dots > \lambda_1 > 0$ . This operation is justified by Equation (53) where it can be seen that the permutation of the eigenvalues does not affect the expectation result. Three cases can be identified from the expressions of KLD.

#### 8.1. Case $1 > \lambda_p > \lambda_{p-1} > \dots > \lambda_1 > 0$

The expression of  $\text{KL}(\mathbf{X}^1 || \mathbf{X}^2)$  is given by Equation (109) and  $\text{KL}(\mathbf{X}^2 || \mathbf{X}^1)$  is given by (115).

#### 8.2. Case $\lambda_p > \lambda_{p-1} > \dots > \lambda_1 > 1$

$\text{KL}(\mathbf{X}^1 || \mathbf{X}^2)$  is given by the Equation (110) and  $\text{KL}(\mathbf{X}^2 || \mathbf{X}^1)$  is given by (114).

#### 8.3. Case $\lambda_p > 1$ and $\lambda_1 < 1$

This case guarantees that  $0 \leq 1 - \lambda_j / \lambda_p < 1, j = 1, \dots, p - 1$  and  $0 \leq 1 - 1 / \lambda_p < 1$ . The expression of the  $\text{KL}(\mathbf{X}^1 || \mathbf{X}^2)$  is given by Equation (106) and  $\text{KL}(\mathbf{X}^2 || \mathbf{X}^1)$  is given by (112) or (113). To perform an evaluation of the quality of the numerical approximation of the derivative of the Lauricella series, we consider a case where an exact and simple expression of  $\left. \frac{\partial}{\partial a} \{ F_D^{(p)}(\cdot) \} \right|_{a=0}$  is possible. The following case where  $\lambda_1 = \dots = \lambda_p = \lambda$  allows the Lauricella series to be equivalent to the Gauss hypergeometric function given as follows

$$F_D^{(p)} \left( a, \underbrace{\frac{1}{2}, \dots, \frac{1}{2}}_p; a + \frac{1+p}{2}; 1 - \lambda, \dots, 1 - \lambda \right) = {}_2F_1 \left( a, \frac{p}{2}; a + \frac{1+p}{2}; 1 - \lambda \right). \tag{132}$$

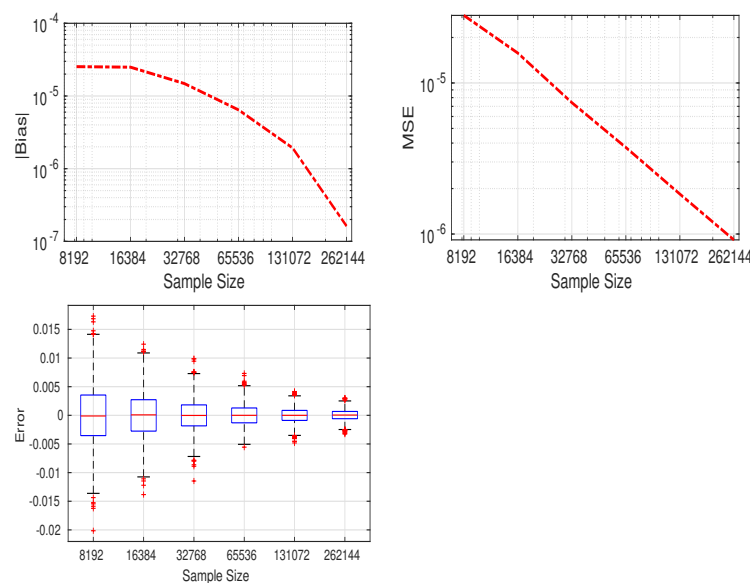
This relation allows us to compare the computational accuracy of the approximation of the Lauricella series with respect to the Gauss function. In addition, to compute the numerical value the indices of the series will evolve from 0 to  $N$  instead of infinity. The latter is chosen to ensure a good approximation of the Lauricella series. Table 2 shows the computation of the derivative of  $F_D^{(p)}(\cdot)$  and  ${}_2F_1(\cdot)$ , along with the absolute value of error  $|\epsilon|$ , where

$p = 2, N = \{20, 30, 40\}$ . The exact expression of  $\frac{\partial}{\partial a}\{ {}_2F_1(\cdot) \}|_{a=0}$  when  $p = 2$  is given by Equation (129). We can deduce the following conclusions. First, the error is reasonably low and decreases as the value of  $N$  increases. Second, the error increases for values of  $1 - \lambda$  close to 1 as expected, which corresponds to the convergence region limit.

**Table 2.** Computation of  $A = \frac{\partial}{\partial a}\{ {}_2F_1(\cdot) \}|_{a=0}$  and  $B = \frac{\partial}{\partial a}\{ F_D^{(p)}(\cdot) \}|_{a=0}$  when  $p = 2$  and  $\lambda_1 = \dots = \lambda_p = \lambda$ .

$1 - \lambda$	$N = 20$			$N = 30$		$N = 40$	
	A	B	$ \epsilon $	B	$ \epsilon $	B	$ \epsilon $
0.1	0.0694	0.0694	$9.1309 \times 10^{-16}$	0.0694	$9.1309 \times 10^{-16}$	0.0694	$9.1309 \times 10^{-16}$
0.3	0.2291	0.2291	$3.7747 \times 10^{-14}$	0.2291	$1.1102 \times 10^{-16}$	0.2291	$1.1102 \times 10^{-16}$
0.5	0.4292	0.4292	$2.6707 \times 10^{-9}$	0.4292	$1.2458 \times 10^{-12}$	0.4292	$6.6613 \times 10^{-16}$
0.7	0.7022	0.7022	$5.9260 \times 10^{-6}$	0.7022	$8.2678 \times 10^{-8}$	0.7022	$1.3911 \times 10^{-9}$
0.9	1.1673	1.1634	0.0038	1.1665	$7.2760 \times 10^{-4}$	1.1671	$1.6081 \times 10^{-4}$
0.99	1.7043	1.5801	0.1241	1.6267	0.0776	1.6514	0.0529

In the following section, we compare the Monte Carlo sampling method to approximate the KLD value with the numerical value of the closed-form expression of the latter. The Monte Carlo method involves sampling a large number of samples and using them to calculate the sum rather than the integral. Here, for each sample size, the experiment is repeated 2000 times. The elements of  $\Sigma_1$  and  $\Sigma_2$  are given in Table 3. Figure 1 depicts the absolute value of bias, mean square error (MSE) and box plot of the difference between the symmetric KL approximated value and its theoretical one, given versus the sample sizes. As the sample size increases, the bias and the MSE decrease. Accordingly, the approximated value will be very close to the theoretical KLD when the number of samples is very large. The computation time of the proposed approximation and the classical Monte Carlo sampling method are recorded using Matlab on a 1.6 GHz processor with 16 GB of memory. For the proposed numerical approximation, the computation time is evaluated to 1.56 s with  $N = 20$ . The value of  $N$  can be increased to further improve the accuracy, but it will increase the computation time. For the Monte Carlo sampling method, the mean time values at sample sizes of {65,536; 131,072; 262,144} are {2.71; 5.46; 10.78} seconds, respectively.



**Figure 1.** Top row: Bias (left) and MSE (right) of the difference between the approximated and theoretical symmetric KL for MCD. Bottom row: Box plot of the error. The mean error is the bias. Outliers are larger than  $Q_3 + 1.5 \times IQR$  or smaller than  $Q_1 - 1.5 \times IQR$ , where  $Q_1$ ,  $Q_3$ , and  $IQR$  are the 25th, 75th percentiles, and the interquartile range, respectively.

**Table 3.** Parameters  $\Sigma_1$  and  $\Sigma_2$  used to compute KLD for central MCD.

$\Sigma$	$\Sigma_{11}, \Sigma_{22}, \Sigma_{33}, \Sigma_{12}, \Sigma_{13}, \Sigma_{23}$
$\Sigma_1$	1, 1, 1, 0.6, 0.2, 0.3
$\Sigma_2$	1, 1, 1, 0.3, 0.1, 0.4

To further encourage the dissemination of these results, we provide a code available as attached file to this paper. This is given in Matlab with a specific case of  $p = 3$ . This can be easily extended to any value of  $p$ , thanks to the general closed-form expression established in this paper.

```

%*****
% Compute the KL divergence and distance between two central multivariate Cauchy
% distribution.
%% Input:
% + Sigma1: Symmetric positive definite (p*p) scale matrix
% + Sigma2: Symmetric positive definite (p*p) scale matrix
% + nb: indices used to compute the KL and dis; nb={20,30,40,etc}.
% Increase nb means increase the precision and also the computation time.
%% Output:
% + KL_12: KL divergence between X1 and X2: KL(X1||X2)
% + KL_21: KL divergence between X2 and X1: KL(X2||X1)
% + Esp_12: expectation E_{X}\{\ln[1+ X^T*Sigma2^{-1}*X]\} where X~MCD(Sigma1,p=3)
% + Esp_21: expectation E_{X}\{\ln[1+ X^T*Sigma1^{-1}*X]\} where X~MCD(Sigma2,p=3)
% + dis: distance between X1 and X2: dis = KL(X1||X2) + KL(X2||X1)
%% Example:
% Sigma1 = [1 0.6 0.2; 0.6 1 0.3; 0.2 0.3 1];
% Sigma2 = [1 0.3 0.1; 0.3 1 0.4; 0.1 0.4 1];
% [KL_12, KL_21, Esp_12, Esp_21, dis] = fonction_KL_MCD_final(Sigma1,Sigma2,20);
%*****
function [KL_12, KL_21, Esp_12, Esp_21, dis] = fonction_KL_MCD_final(Sigma1,Sigma2,nb)
format long;
p = 3;

vpr = real(eig(Sigma1*inv(Sigma2)));
vpr = sort(vpr,'ascend');

nbre = nb;
[N,M,L] = ndgrid(0:nbre,0:nbre,0:nbre);

if vpr(p)< 1
%*****
%% Derivative of Fd(a,1/2,1/2,1/2;a+(1+p)/2;1-vpr(1),1-vpr(2),1-vpr(3)) | a=0
%*****
H = N+M+L;
H(H==0)= inf;
commun = (1-vpr(1)).^N./ factorial(N).*(1-vpr(2)).^M./ factorial(M).*(1-vpr(3)).^L./ factorial(L).*...
    pochhammer(1/2,N).* pochhammer(1/2,M).* pochhammer(1/2,L);
h1 = commun.* pochhammer(1,N+M+L)./ pochhammer((1+p)/2, N+M+L)*1./H;
derive1 = sum(sum(sum(h1))) % Eq. (102) and (A1)
%-----
J = N+M+L-1;
A=[0, cumsum(1./((p+1)/2 +(0:p*nbre-1)))]';
for i=1:nbre+1
    for j=1:nbre+1
        for l=1:nbre+1
            G(i,j,l)= -A(J(i,j,l)+2);
        end
    end
end
end
h2 = commun.*G;
derive2 = sum(sum(sum(h2))) % Eq. (104) and (A1)
%-----
Esp_12 = psi(1/2 + p/2)-psi(1/2) - derive1
Esp_21 = psi(1/2 + p/2)-psi(1/2) - prod(vpr).^(1/2)* derive2
KL_12 = -1/2*log(prod(vpr))-(1+p)/2*derive1 % Eq.(109)

```

```

KL_21 = 1/2*log(prod(vpr))-(1+p)/2*prod(vpr).^(1/2)*derive2 % Eq.(115)

elseif vpr(1) > 1
%*****
%% Derivative of Fd((1+p)/2,1/2,1/2,1/2;a+(1+p)/2;1-1/vpr(1),1-1/vpr(2),1-1/vpr(3)) | a=0
%*****
J = N+M+L-1;
A=[0, cumsum(1./((p+1)/2 +(0:p*nbre-1)))];
for i=1:nbre+1
    for j=1:nbre+1
        for l=1:nbre+1
            G(i,j,l)= -A(J(i,j,l)+2);
        end
    end
end
commun = (1-1/vpr(1)).^N./ factorial(N).*(1-1/vpr(2)).^M./ factorial(M).*(1-1/vpr(3)).^L./ factorial(L).*...
    pochhammer(1/2,N).* pochhammer(1/2,M).* pochhammer(1/2,L);
h1 = commun.*G;
derive1 = sum(sum(sum(h1))) % Eq. (104) and (A1)
%-----
H = N+M+L;
H(H=0)= inf;
h2 = commun.* pochhammer(1,N+M+L)./ pochhammer((1+p)/2, N+M+L)*1./H;
derive2 = sum(sum(sum(h2))) % Eq. (102) and (A1)
%-----
Esp_12 = psi(1/2 + p/2)-psi(1/2) - prod(vpr).^(-1/2)*derive1
Esp_21 = psi(1/2 + p/2)-psi(1/2) - derive2
KL_12 = -1/2*log(prod(vpr))-(1+p)/2*prod(vpr).^(-1/2)*derive1 % Eq.(110)
KL_21 = 1/2*log(prod(vpr))-(1+p)/2*derive2 % Eq.(114)

else
%*****
%% Derivative of Fd(a,1/2,1/2,a+1/2;a+(1+p)/2;1-vpr(1)/vpr(3),1-vpr(2)/vpr(3),1-1/vpr(3)) | a=0
%*****
H = N+M+L;
H(H=0)= inf;
commun = (1-vpr(1)/vpr(3)).^N./ factorial(N).*(1-vpr(2)/vpr(3)).^M./ factorial(M).*...
    (1-1/vpr(3)).^L./ factorial(L).*...
    pochhammer(1/2,N).* pochhammer(1/2,M).* pochhammer(1/2,L);
h1 = commun.* pochhammer(1,N+M+L)./ pochhammer((1+p)/2,N+M+L)./H;
derive1 = sum(sum(sum(h1))) % Eq.(101) and (A1)
%-----
J = N+M+L-1;
JJ = L-1;
A = [0, cumsum(1./((p+1)/2 +(0:p*nbre-1)))];
B = [0, cumsum(1./(1/2 +(0:nbre-1)))];
for i=1:nbre+1
    for j=1:nbre+1
        for l=1:nbre+1
            G(i,j,l)= B(JJ(i,j,l)+2) - A(J(i,j,l)+2);
        end
    end
end
h2 = commun.*G;
derive2 = sum(sum(sum(h2))) % Eq. (103) and (A1)
%-----
Esp_12 = psi(1/2 + p/2)-psi(1/2) + log(vpr(p)) - derive1
Esp_21 = psi(1/2 + p/2)-psi(1/2) - vpr(p)^(-p/2)*(vpr(1)*vpr(2))^(1/2)*derive2
KL_12 = -1/2*log(prod(vpr)) -(1+p)/2*(-log(vpr(p)) + derive1) % Eq. (106)
KL_21 = 1/2*log(prod(vpr)) -(1+p)/2*vpr(p)^(-p/2)*(vpr(1)*vpr(2))^(1/2)*derive2 % Eq. (112)
end

dis = KL_12 + KL_21

```

## 9. Conclusions

Since the MCDs have various applications in signal and image processing, the KLD between central MCDs tackles an important problem for future work on statistics, machine learning and other related fields in computer science. In this paper, we derived a closed-

form expression of the KLD and distance between two central MCDs. The similarity measure can be expressed as function of the Lauricella D-hypergeometric series  $F_D^{(p)}$ . We have also proposed a simple scheme to compute easily the Lauricella series and to bypass the convergence constraints of this series. Codes and examples for numerical calculations are presented and explained in detail. Finally, a comparison is made to show how the Monte Carlo sampling method gives approximations close to the KLD theoretical value. As a final note, it is also possible to extend these results on the KLD to the case of the multivariate  $t$ -distribution since the MCD is a particular case of this multivariate distribution.

**Author Contributions:** Conceptualization, N.B.; methodology, N.B.; software, N.B.; writing original draft preparation, N.B.; writing review and editing, N.B. and D.R.; supervision, D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Authors gratefully acknowledge the PHENOTIC platform node of the french national infrastructure on plant phenotyping ANR PHENOME 11-INBS-0012. The authors would like also to thank the anonymous reviewers for their helpful comments valuable comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Lauricella Function

In 1893, G. Lauricella [39] investigated the properties of four series  $F_A^{(n)}, F_B^{(n)}, F_C^{(n)}, F_D^{(n)}$  of  $n$  variables. When  $n = 2$ , these functions coincide with Appell's  $F_2, F_3, F_4, F_1$ , respectively. When  $n = 1$ , they all coincide with Gauss'  ${}_2F_1$ . We present here only the Lauricella series  $F_D^{(n)}$  given as follows

$$F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) = \sum_{m_1=0}^{\infty} \dots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\dots+m_n} (b_1)_{m_1} \dots (b_n)_{m_n} x_1^{m_1} \dots x_n^{m_n}}{(c)_{m_1+\dots+m_n} m_1! \dots m_n!} \tag{A1}$$

where  $|x_1|, \dots, |x_n| < 1$ . The Pochhammer symbol  $(q)_i$  indicates the  $i$ -th rising factorial of  $q$ , i.e.,

$$(q)_i = q(q+1) \dots (q+i-1) = \frac{\Gamma(q+i)}{\Gamma(q)} \quad \text{if } i = 1, 2, \dots \tag{A2}$$

If  $i = 0$ ,  $(q)_i = 1$ . Function  $F_D^{(n)}(\cdot)$  can be expressed in terms of multiple integrals as follows [42]

$$F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) = \frac{\Gamma(c)}{\Gamma(c - \sum_{i=1}^n b_i) \prod_{i=1}^n \Gamma(b_i)} \times \int_{\Omega} \dots \int \prod_{i=1}^n u_i^{b_i-1} (1 - \sum_{i=1}^n u_i)^{c - \sum_{i=1}^n b_i - 1} (1 - \sum_{i=1}^n x_i u_i)^{-a} \prod_{i=1}^n du_i \tag{A3}$$

where  $\Omega = \{(u_1, u_2, \dots, u_n); 0 \leq u_i \leq 1, i = 1, \dots, n, \text{ and } 0 \leq u_1 + u_2 + \dots + u_n \leq 1\}$ ,  $\text{Real}(b_i) > 0$  for  $i = 1, \dots, n$  and  $\text{Real}(c - b_1 - \dots - b_n) > 0$ . Lauricella's  $F_D$  can be written as a one-dimensional Euler-type integral for any number  $n$  of variables. The integral form of  $F_D^{(n)}(\cdot)$  is given as follows when  $\text{Real}(a) > 0$  and  $\text{Real}(c - a) > 0$

$$F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{a-1} (1-u)^{c-a-1} (1-ux_1)^{-b_1} \dots (1-ux_n)^{-b_n} du. \tag{A4}$$

Lauricella has given several transformation formulas, from which we use the two following relationships. More details can be found in Exton’s book [43] on hypergeometric equations.

$$F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n) = \prod_{i=1}^n (1 - x_i)^{-b_i} F_D^{(n)}\left(c - a, b_1, \dots, b_n; c; \frac{x_1}{x_1 - 1}, \dots, \frac{x_n}{x_n - 1}\right) \tag{A5}$$

$$= (1 - x_1)^{-a} F_D^{(n)}\left(a, c - \sum_{i=1}^n b_i, b_2, \dots, b_n; c; \frac{x_1}{x_1 - 1}, \frac{x_1 - x_2}{x_1 - 1}, \dots, \frac{x_1 - x_n}{x_1 - 1}\right) \tag{A6}$$

$$= (1 - x_n)^{-a} F_D^{(n)}\left(a, b_1, \dots, b_{n-1}, c - \sum_{i=1}^n b_i; c; \frac{x_n - x_1}{x_n - 1}, \frac{x_n - x_2}{x_n - 1}, \dots, \frac{x_n - x_{n-1}}{x_n - 1}, \frac{x_n}{x_n - 1}\right) \tag{A7}$$

$$= (1 - x_1)^{c-a} \prod_{i=1}^n (1 - x_i)^{-b_i} F_D^{(n)}\left(c - a, c - \sum_{i=1}^n b_i, b_2, \dots, b_n; c; x_1, \frac{x_2 - x_1}{x_2 - 1}, \dots, \frac{x_n - x_1}{x_n - 1}\right) \tag{A8}$$

$$= (1 - x_n)^{c-a} \prod_{i=1}^n (1 - x_i)^{-b_i} F_D^{(n)}\left(c - a, b_1, \dots, b_{n-1}, c - \sum_{i=1}^n b_i; c; \frac{x_1 - x_n}{x_1 - 1}, \dots, \frac{x_{n-1} - x_n}{x_{n-1} - 1}, x_n\right). \tag{A9}$$

**Appendix B. Demonstration of Derivative**

*Appendix B.1. Demonstration*

We use the following notation  $\alpha = \sum_{i=1}^p m_i$  to alleviate the writing of equations. Knowing that  $\frac{\partial}{\partial c}(c)_k = (c)_k(\psi(c + k) - \psi(c))$ ,  $\psi(c + k) - \psi(c) = \sum_{\ell=0}^{k-1} \frac{1}{c+\ell}$  and  $(c)_k = \prod_{i=0}^{k-1} (c + i)$  we can state that

$$\begin{aligned} \frac{\partial}{\partial a} \left\{ \frac{(a)_\alpha}{(a + \frac{1+p}{2})_\alpha} \right\} &= \frac{(a)_\alpha [\psi(a + \alpha) - \psi(a) - \psi(a + \frac{1+p}{2} + \alpha) + \psi(a + \frac{1+p}{2})]}{(a + \frac{1+p}{2})_\alpha} \\ &= \frac{\prod_{k=0}^{\alpha-1} (a + k) \left( \sum_{k=0}^{\alpha-1} \frac{1}{a+k} - \frac{1}{a + \frac{1+p}{2} + k} \right)}{(a + \frac{1+p}{2})_\alpha}. \end{aligned} \tag{A10}$$

Using the fact that

$$\prod_{k=0}^{\alpha-1} (a + k) \sum_{k=0}^{\alpha-1} \frac{1}{a + k} = \prod_{k=1}^{\alpha-1} (a + k) + \prod_{k=0, k \neq 1}^{\alpha-1} (a + k) + \dots + \prod_{k=0}^{\alpha-2} (a + k) \tag{A11}$$

we can state that

$$\left. \frac{\partial}{\partial a} \left\{ \frac{(a)_\alpha}{(a + \frac{1+p}{2})_\alpha} \right\} \right|_{a=0} = \frac{(\alpha - 1)!}{(\frac{1+p}{2})_\alpha} = \frac{(1)_\alpha}{(\frac{1+p}{2})_\alpha} \frac{1}{\alpha}. \tag{A12}$$

*Appendix B.2. Demonstration*

$$\begin{aligned} \frac{\partial}{\partial a} \left\{ \frac{(a)_\alpha (a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_\alpha} \right\} &= \frac{(a + \frac{1}{2})_{m_p} (a)_\alpha [\psi(a + \alpha) - \psi(a) + \psi(a + \frac{1}{2} + m_p) - \psi(a + \frac{1}{2})]}{(a + \frac{1+p}{2})_\alpha} \\ &\quad - \frac{(a)_\alpha (a + \frac{1}{2})_{m_p} [\psi(a + \frac{1+p}{2} + \alpha) - \psi(a + \frac{1+p}{2})]}{(a + \frac{1+p}{2})_\alpha} \end{aligned} \tag{A13}$$

$$= \frac{(a + \frac{1}{2})_{m_p} \prod_{k=0}^{\alpha-1} (a + k) \left[ \sum_{k=0}^{\alpha-1} \frac{1}{a+k} - \frac{1}{a + \frac{1+p}{2} + k} + \sum_{k=0}^{m_p-1} \frac{1}{a + \frac{1}{2} + k} \right]}{(a + \frac{1+p}{2})_\alpha}. \tag{A14}$$



By developing the previous expression we can state that

$$\frac{\partial}{\partial a} \left\{ \frac{(a)_\alpha (a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_\alpha} \right\} \Big|_{a=0} = \frac{(\frac{1}{2})_{m_p} (\alpha - 1)!}{(\frac{1+p}{2})_\alpha} = \frac{(\frac{1}{2})_{m_p} (1)_\alpha}{(\frac{1+p}{2})_\alpha} \frac{1}{\alpha}. \tag{A15}$$

Appendix B.3. Demonstration

$$\frac{\partial}{\partial a} \left\{ \frac{(a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_\alpha} \right\} = \frac{(a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_\alpha} \left[ \sum_{k=0}^{m_p-1} \frac{1}{a + \frac{1}{2} + k} - \sum_{k=0}^{\alpha-1} \frac{1}{a + \frac{1+p}{2} + k} \right]. \tag{A16}$$

As a consequence,

$$\frac{\partial}{\partial a} \left\{ \frac{(a + \frac{1}{2})_{m_p}}{(a + \frac{1+p}{2})_\alpha} \right\} \Big|_{a=0} = \frac{(\frac{1}{2})_{m_p}}{(\frac{1+p}{2})_\alpha} \left[ \sum_{k=0}^{m_p-1} \frac{1}{\frac{1}{2} + k} - \sum_{k=0}^{\alpha-1} \frac{1}{\frac{1+p}{2} + k} \right]. \tag{A17}$$

Appendix B.4. Demonstration

$$\frac{\partial}{\partial a} \left\{ \frac{1}{(a + \frac{1+p}{2})_\alpha} \right\} = - \frac{\psi(a + \frac{1+p}{2} + \alpha) - \psi(a + \frac{1+p}{2})}{(a + \frac{1+p}{2})_\alpha} \tag{A18}$$

$$= \frac{-1}{(a + \frac{1+p}{2})_\alpha} \sum_{k=0}^{\alpha-1} \frac{1}{a + \frac{1+p}{2} + k}. \tag{A19}$$

Finally,

$$\frac{\partial}{\partial a} \left\{ \frac{1}{(a + \frac{1+p}{2})_\alpha} \right\} \Big|_{a=0} = \frac{-1}{(\frac{1+p}{2})_\alpha} \sum_{k=0}^{\alpha-1} \frac{1}{\frac{1+p}{2} + k}. \tag{A20}$$

Appendix C. Computations of Some Equations

Appendix C.1. Computation

Let  $f$  be a function of  $\lambda$  defined as follows:

$$f(\lambda) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)_n \frac{1}{n} \frac{(1-\lambda)^n}{n!}. \tag{A21}$$

The multiplication of the derivative of  $f$  with respect to  $\lambda$  by  $(1 - \lambda)$  is given as follows

$$(1 - \lambda) \frac{\partial}{\partial \lambda} f(\lambda) = - \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)_n \frac{(1-\lambda)^n}{n!} = 1 - \lambda^{-1/2}. \tag{A22}$$

As a consequence,

$$\frac{\partial}{\partial \lambda} f(\lambda) = \frac{1 - \lambda^{-1/2}}{1 - \lambda} = \frac{-\lambda^{-1/2}}{1 + \lambda^{1/2}}. \tag{A23}$$

Finally,

$$f(\lambda) = -2 \ln \frac{1 + \lambda^{1/2}}{2}. \tag{A24}$$

Appendix C.2. Computation

$$\frac{\partial}{\partial a} \left\{ {}_2F_1 \left( a, \frac{1}{2}; a + \frac{3}{2}; 1 - \lambda_i \right) \right\} \Big|_{a=0} = \sum_{n=1}^{\infty} \frac{(\frac{1}{2})_n (1)_n}{(\frac{3}{2})_n n} \frac{(1 - \lambda_i)^n}{n!} = f(\lambda_i) \tag{A25}$$

where  $f$  is a function of  $\lambda_i$ . The multiplication of the derivative of  $f$  with respect to  $\lambda_i$  by  $(1 - \lambda_i)$  is given as follows

$$(1 - \lambda_i) \frac{\partial}{\partial \lambda_i} f(\lambda_i) = - \sum_{n=1}^{\infty} \frac{(\frac{1}{2})_n (1)_n}{(\frac{3}{2})_n} \frac{(1 - \lambda_i)^n}{n!} \tag{A26}$$

$$= - {}_2F_1 \left( \frac{1}{2}, 1; \frac{3}{2}; 1 - \lambda_i \right) + 1. \tag{A27}$$

Knowing that

$${}_2F_1 \left( \frac{1}{2}, 1; \frac{3}{2}; 1 - \lambda_i \right) = \frac{\arctan(\sqrt{\lambda_i - 1})}{\sqrt{\lambda_i - 1}} \tag{A28}$$

$$= \frac{1}{2\sqrt{1 - \lambda_i}} \ln \left( \frac{1 + \sqrt{1 - \lambda_i}}{1 - \sqrt{1 - \lambda_i}} \right) \tag{A29}$$

we can deduce an expression of

$$\frac{\partial}{\partial \lambda_i} f(\lambda_i) = \frac{\arctan(\sqrt{\lambda_i - 1})}{(\lambda_i - 1)^{3/2}} + \frac{1}{1 - \lambda_i}. \tag{A30}$$

Accordingly,

$$f(\lambda_i) = - \ln \lambda_i - 2 \frac{\arctan(\sqrt{\lambda_i - 1})}{\sqrt{\lambda_i - 1}} + 2 \tag{A31}$$

$$= - \ln \lambda_i + \frac{1}{\sqrt{1 - \lambda_i}} \ln \left( \frac{1 - \sqrt{1 - \lambda_i}}{1 + \sqrt{1 - \lambda_i}} \right) + 2. \tag{A32}$$

Appendix C.3. Computation

$$\frac{\partial}{\partial a} \left\{ {}_2F_1 \left( a, 1; a + \frac{3}{2}; 1 - \lambda \right) \right\} \Big|_{a=0} = \sum_{n=1}^{\infty} \frac{(1)_n (1)_n}{(\frac{3}{2})_n n} \frac{(1 - \lambda)^n}{n!} = f(\lambda) \tag{A33}$$

where  $f$  is a function of  $\lambda$ . The multiplication of the derivative of  $f$  with respect to  $\lambda$  by  $(1 - \lambda)$  is given as follows

$$(1 - \lambda) \frac{\partial}{\partial \lambda} f(\lambda) = - \sum_{n=1}^{\infty} \frac{(1)_n (1)_n}{(\frac{3}{2})_n} \frac{(1 - \lambda)^n}{n!} \tag{A34}$$

$$= - {}_2F_1 \left( 1, 1; \frac{3}{2}; 1 - \lambda \right) + 1. \tag{A35}$$

Knowing that

$${}_2F_1 \left( 1, 1; \frac{3}{2}; 1 - \lambda \right) = \frac{1}{\sqrt{\lambda}} \frac{\arcsin(\sqrt{1 - \lambda})}{\sqrt{1 - \lambda}} \tag{A36}$$

we can state that

$$\frac{\partial}{\partial \lambda} f(\lambda) = - \frac{1}{\sqrt{\lambda}} \frac{\arcsin(\sqrt{1 - \lambda})}{(1 - \lambda)^{3/2}} + \frac{1}{1 - \lambda}. \tag{A37}$$

As a consequence,

$$f(\lambda) = -\frac{2\sqrt{\lambda} \arcsin(\sqrt{1-\lambda})}{\sqrt{1-\lambda}} + 2 \quad (\text{A38})$$

$$= -\frac{2}{\sqrt{1-\lambda^{-1}}} \ln(\sqrt{\lambda} + \sqrt{\lambda-1}) + 2. \quad (\text{A39})$$

## References

- Ollila, E.; Tyler, D.E.; Koivunen, V.; Poor, H.V. Complex Elliptically Symmetric Distributions: Survey, New Results and Applications. *IEEE Trans. Signal Process.* **2012**, *60*, 5597–5625. [\[CrossRef\]](#)
- Kotz, S.; Nadarajah, S. *Multivariate T-Distributions and Their Applications*; Cambridge University Press: Cambridge, UK, 2004. [\[CrossRef\]](#)
- Press, S. Multivariate stable distributions. *J. Multivar. Anal.* **1972**, *2*, 444–462. [\[CrossRef\]](#)
- Sahu, S.; Singh, H.V.; Kumar, B.; Singh, A.K. Statistical modeling and Gaussianization procedure based de-speckling algorithm for retinal OCT images. *J. Ambient. Intell. Humaniz. Comput.* **2018**, 1–14. [\[CrossRef\]](#)
- Ranjani, J.J.; Thiruvengadam, S.J. Generalized SAR Despeckling Based on DTCWT Exploiting Interscale and Intrascale Dependences. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 552–556. [\[CrossRef\]](#)
- Sadrezami, H.; Ahmad, M.O.; Swamy, M.N.S. Color image denoising using multivariate cauchy PDF in the contourlet domain. In Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 15–18 May 2016; pp. 1–4. [\[CrossRef\]](#)
- Sadrezami, H.; Ahmad, M.O.; Swamy, M.N.S. A Study of Multiplicative Watermark Detection in the Contourlet Domain Using Alpha-Stable Distributions. *IEEE Trans. Image Process.* **2014**, *23*, 4348–4360. [\[CrossRef\]](#)
- Fontaine, M.; Nugraha, A.A.; Badeau, R.; Yoshii, K.; Liutkus, A. Cauchy Multichannel Speech Enhancement with a Deep Speech Prior. In Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2–6 September 2019; pp. 1–5.
- Cover, T.M.; Thomas, J.A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*; Wiley-Interscience: Hoboken, NJ, USA, 2006. [\[CrossRef\]](#)
- Pardo, L. *Statistical Inference Based on Divergence Measures*; CRC Press: Abingdon, UK, 2005.
- Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [\[CrossRef\]](#)
- Kullback, S. *Information Theory and Statistics*; Wiley: New York, NY, USA, 1959.
- Rényi, A. On Measures of Entropy and Information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
- Sharma, B.D.; Mittal, D.P. New non-additive measures of relative information. *J. Comb. Inf. Syst. Sci.* **1977**, *2*, 122–132.
- Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **1943**, *35*, 99–109.
- Kailath, T. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Trans. Commun. Technol.* **1967**, *15*, 52–60. [\[CrossRef\]](#)
- Giet, L.; Lubrano, M. A minimum Hellinger distance estimator for stochastic differential equations: An application to statistical inference for continuous time interest rate models. *Comput. Stat. Data Anal.* **2008**, *52*, 2945–2965. [\[CrossRef\]](#)
- Csiszár, I. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **1963**, *8*, 85–108.
- Ali, S.M.; Silvey, S.D. A General Class of Coefficients of Divergence of One Distribution from Another. *J. R. Stat. Soc. Ser. B (Methodol.)* **1966**, *28*, 131–142. [\[CrossRef\]](#)
- Bregman, L. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217. [\[CrossRef\]](#)
- Burbea, J.; Rao, C. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 489–495. [\[CrossRef\]](#)
- Burbea, J.; Rao, C. On the convexity of higher order Jensen differences based on entropy functions (Corresp.). *IEEE Trans. Inf. Theory* **1982**, *28*, 961–963. [\[CrossRef\]](#)
- Burbea, J.; Rao, C. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multivar. Anal.* **1982**, *12*, 575–596. [\[CrossRef\]](#)
- Csiszar, I. Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* **1967**, *2*, 229–318.
- Nielsen, F.; Nock, R. On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Process. Lett.* **2014**, *21*, 10–13. [\[CrossRef\]](#)
- Menéndez, M.L.; Morales, D.; Pardo, L.; Salicrú, M. Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: A unified study. *Stat. Pap.* **1995**, *36*, 1–29. [\[CrossRef\]](#)
- Cover, T.M.; Thomas, J.A. Information theory and statistics. *Elem. Inf. Theory* **1991**, *1*, 279–335.

28. MacKay, D.J.C. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
29. Ruiz, F.E.; Pérez, P.S.; Bonev, B.I. *Information Theory in Computer Vision and Pattern Recognition*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
30. Nielsen, F. Statistical Divergences between Densities of Truncated Exponential Families with Nested Supports: Duo Bregman and Duo Jensen Divergences. *Entropy* **2022**, *24*, 421. [[CrossRef](#)] [[PubMed](#)]
31. Chyzak, F.; Nielsen, F. A closed-form formula for the Kullback–Leibler divergence between Cauchy distributions. *arXiv* **2019**, arXiv:1905.10965.
32. Nielsen, F.; Okamura, K. On f-divergences between Cauchy distributions. *arXiv* **2021**, arXiv:2101.12459.
33. Srivastava, H.; Karlsson, P.W. *Multiple Gaussian Hypergeometric Series*; Ellis Horwood Series in Mathematics and Its Applications Statistics and Operational Research, E; Horwood Halsted Press: Chichester, UK; West Sussex, UK; New York, NY, USA, 1985.
34. Mathai, A.M.; Haubold, H.J. *Special Functions for Applied Scientists*; Springer Science+Business Media: New York, NY, USA, 2008.
35. Gradshteyn, I.; Ryzhik, I. *Table of Integrals, Series, and Products*, 7th ed.; Academic Press is an Imprint of Elsevier: Cambridge, MA, USA, 2007.
36. Humbert, P. The Confluent Hypergeometric Functions of Two Variables. *Proc. R. Soc. Edinb.* **1922**, *41*, 73–96. [[CrossRef](#)]
37. Erdélyi, A. *Higher Transcendental Functions*; McGraw-Hill: New York, NY, USA, 1953; Volume I.
38. Koepf, W. *Hypergeometric Summation an Algorithmic Approach to Summation and Special Function Identities*, 2nd ed.; Universitext, Springer: London, UK, 2014.
39. Lauricella, G. Sulle funzioni ipergeometriche a piu variabili. *Rend. Del Circ. Mat. Palermo* **1893**, *7*, 111–158. [[CrossRef](#)]
40. Mathai, A.M. *Jacobians of Matrix Transformations and Functions of Matrix Argument*; World Scientific: Singapore, 1997.
41. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
42. Hattori, A.; Kimura, T. On the Euler integral representations of hypergeometric functions in several variables. *J. Math. Soc. Jpn.* **1974**, *26*, 1–16. [[CrossRef](#)]
43. Exton, H. *Multiple Hypergeometric Functions and Applications*; Wiley: New York, NY, USA, 1976.