entropy

MDPI

Article

Exact Expressions for Kullback–Leibler Divergence for Multivariate and Matrix-Variate Distributions

Victor Nawa [1] and Saralees Nadarajah [2,*]

[1] Department of Mathematics and Statistics, University of Zambia, Lusaka 10101, Zambia; vnawa@yahoo.com
[2] Department of Mathematics, University of Manchester, Manchester M13 9PL, UK
* Correspondence: mbbsssn2@manchester.ac.uk

Abstract: The Kullback–Leibler divergence is a measure of the divergence between two probability distributions, often used in statistics and information theory. However, exact expressions for it are not known for multivariate or matrix-variate distributions apart from a few cases. In this paper, exact expressions for the Kullback–Leibler divergence are derived for over twenty multivariate and matrix-variate distributions. The expressions involve various special functions.

Keywords: beta function; gamma function; matrix-variate beta function; matrix-variate confluent hypergeometric function; matrix-variate gamma function; matrix-variate Gauss hypergeometric function; matrix-variate type I Dirichlet density; type I Dirichlet density

1. Introduction

The Kullback–Leibler divergence (KLD) due to [1] is a fundamental concept in information theory and statistics used to measure the divergence between two probability distributions. It quantifies how one probability distribution diverges from a second, reference probability distribution. Specifically, it calculates the expected extra amount of information required to represent data sampled from one distribution using a code optimized for another distribution. The KLD is asymmetric and not a true metric as it does not satisfy the triangle inequality. It is widely employed in various fields including machine learning, where it serves as a key component in tasks such as model comparison, optimization, and generative modeling, providing a measure of dissimilarity or discrepancy between probability distributions [2].

Suppose $\mathbf{X}$ is a continuous vector-variate random variable or a continuous matrix-variate random variable having one of two probability density functions $f_i(\cdot;\boldsymbol{\theta}_i)$, $i=1,2$ parameterized by $\boldsymbol{\theta}_i$, $i=1,2$. The KLD between $f_1(\cdot;\boldsymbol{\theta}_1)$ and $f_2(\cdot;\boldsymbol{\theta}_2)$ is defined by

$$KLD = E\left[\log \frac{f_1(\mathbf{X};\boldsymbol{\theta}_1)}{f_2(\mathbf{X};\boldsymbol{\theta}_2)}\right], \tag{1}$$

where the expectation is with respect to $f_1(\cdot;\boldsymbol{\theta}_1)$.

Because of the increasing applications of the KLD, it is useful to have exact expressions for (1). Apart from the multivariate normal distribution, not many expressions have been derived for (1) for multivariate or matrix-variate distributions. The KLD for the multivariate generalized Gaussian distribution was derived only in 2019, see [3]. The KLD for the multivariate Cauchy distribution was derived only in 2022, see [4]. The KLD for the multivariate t distribution was derived only in 2023, see [5].

The aim of this paper is to derive exact expressions for (1) for over twenty multivariate and matrix-variate distributions. The exact expressions for multivariate distributions are presented in Section 2. The exact expressions for matrix-variate distributions are presented in Section 3. The derivations of all of the expressions including a technical lemma needed

check for updates

Citation: Nawa, V.; Nadarajah, S. Exact Expressions for Kullback–Leibler Divergence for Multivariate and Matrix-Variate Distributions. *Entropy* **2024**, *26*, 663. https://doi.org/10.3390/e26080663

Academic Editor: Lu Wei

Received: 9 July 2024
Revised: 2 August 2024
Accepted: 3 August 2024
Published: 4 August 2024

for the derivations are presented in Section 4. The distributions considered in this paper are continuous. We shall not be considering discrete distributions including mixtures.

The functions and parameters used in this paper are all real-valued. The calculations involve several real-valued special functions listed in the Appendix A.

2. Exact Expressions for Multivariate Distributions

In this section, we state the exact expressions for (1) for Dirichlet, multivariate generalized Gaussian, inverted Dirichlet, multivariate Gauss hypergeometric, multivariate Kotz type, ref. [6]'s multivariate logistic, ref. [7]'s multivariate logistic, ref. [8]'s multivariate normal, multivariate Pearson type II, multivariate Selberg beta, multivariate weighted exponential and von Mises distributions.

A closed form for (1) for the multivariate generalized Gaussian distribution was derived by [3]. But it involved a special function defined as a $(p-1)$ folded infinite sum. The expression we give in Section 2.2 is much simpler in that it involves a single infinite sum. A closed form for (1) for the Dirichlet distribution is available in [9] and https://statproofbook.github.io/P/dir-kl.html (accessed on 1 July 2024).

2.1. Dirichlet Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{\prod_{i=1}^{K} x_i^{a_i-1}}{B(a_1,\ldots,a_{K-1};a_K)}$$

and

$$f_2(\mathbf{x}) = \frac{\prod_{i=1}^{K} x_i^{b_i-1}}{B(b_1,\ldots,b_{K-1};b_K)}$$

for $K \geq 2$, $a_1 > 0,\ldots,a_K > 0$, $b_1 > 0,\ldots,b_K > 0$, $0 \leq x_1 \leq 1,\ldots,0 \leq x_K \leq 1$ and $x_1+\cdots+x_K = 1$. The corresponding KLD is

$$KLD = \log\left[\frac{B(b_1,\ldots,b_{K-1};b_K)}{B(a_1,\ldots,a_{K-1};a_K)}\right] + \sum_{i=1}^{K}(a_i-b_i)\psi(a_i) - \psi(a_1+\cdots+a_K)\sum_{i=1}^{K}(a_i-b_i).$$

2.2. Multivariate Generalized Gaussian Distribution ([10], p. 215)

Consider the joint probability density functions

$$f_1(x_1,\ldots,x_p) = \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}|\mathbf{V}_1|^{-\frac{1}{2}}\exp\left\{-\left(\mathbf{x}^T\mathbf{V}_1^{-1}\mathbf{x}\right)^{\frac{\alpha}{2}}\right\}$$

and

$$f_2(x_1,\ldots,x_p) = \frac{\beta\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\beta}\right)}|\mathbf{V}_2|^{-\frac{1}{2}}\exp\left\{-\left(\mathbf{x}^T\mathbf{V}_2^{-1}\mathbf{x}\right)^{\frac{\beta}{2}}\right\}$$

for $-\infty < x_1 < \infty,\ldots,-\infty < x_p < \infty$, $\alpha > 0$, $\beta > 0$ and $\mathbf{V}_1$, $\mathbf{V}_2$ positive definite symmetric matrices. The corresponding KLD is

$$KLD = \log\left[\frac{\alpha\Gamma\left(\frac{p}{\beta}\right)}{\beta\Gamma\left(\frac{p}{\alpha}\right)}\right] + \frac{1}{2}\log\frac{|\mathbf{V}_2|}{|\mathbf{V}_1|} - \frac{p}{\alpha} + \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{p+\beta}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{\frac{\beta}{2}}{r_1}$$
$$\times\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}\lambda_1^{\frac{\beta}{2}-r_1}\prod_{j=1}^{p-1}\left[\left(\lambda_{j+1}-\lambda_j\right)^{r_j-r_{j+1}}B\left(r_j+\frac{p-j}{2},\frac{1}{2}\right)\right]$$

provided that the infinite sum converges.

2.3. Inverted Dirichlet Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{\Gamma(a_1 + \cdots + a_{K+1})}{\Gamma(a_1) \cdots \Gamma(a_{K+1})} \left(1 + \sum_{i=1}^{K} x_i\right)^{-a_1 - \cdots - a_{K+1}} \prod_{i=1}^{K} x_i^{a_i - 1}$$

and

$$f_2(\mathbf{x}) = \frac{\Gamma(b_1 + \cdots + b_{K+1})}{\Gamma(b_1) \cdots \Gamma(b_{K+1})} \left(1 + \sum_{i=1}^{K} x_i\right)^{-b_1 - \cdots - b_{K+1}} \prod_{i=1}^{K} x_i^{b_i - 1}$$

for $K \geq 2$, $a_1 > 0, \ldots, a_{K+1} > 0$, $b_1 > 0, \ldots, b_{K+1} > 0$ and $x_1 > 0, \ldots, x_K > 0$. The corresponding KLD is

$$\begin{aligned} KLD = {} & \log\left[\frac{\Gamma(a_1 + \cdots + a_{K+1})\Gamma(b_1) \cdots \Gamma(b_{K+1})}{\Gamma(b_1 + \cdots + b_{K+1})\Gamma(a_1) \cdots \Gamma(a_{K+1})}\right] + \sum_{i=1}^{K} (a_i - b_i)\psi(a_i) \\ & - \psi(a_{K+1}) \sum_{i=1}^{K} (a_i - b_i) \\ & + (b_1 + \cdots + b_{K+1} - a_1 - \cdots - a_{K+1})[\psi(a_1 + \cdots + a_{K+1}) - \psi(a_{K+1})]. \end{aligned}$$

2.4. Multivariate Gauss Hypergeometric Distribution [11]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = C(a_1, \ldots, a_K, b, c) \frac{\left(1 - \sum_{i=1}^{K} x_i\right)^{b-1} \prod_{i=1}^{K} x_i^{a_i - 1}}{\left(1 + \sum_{i=1}^{K} x_i\right)^{c}}$$

and

$$f_2(\mathbf{x}) = C(d_1, \ldots, d_K, e, f) \frac{\left(1 - \sum_{i=1}^{K} x_i\right)^{e-1} \prod_{i=1}^{K} x_i^{d_i - 1}}{\left(1 + \sum_{i=1}^{K} x_i\right)^{f}}$$

for $K \geq 2$, $a_1 > 0, \ldots, a_K > 0$, $b > 0$, $-\infty < c < \infty$, $d_1 > 0, \ldots, d_K > 0$, $e > 0$, $-\infty < f < \infty$, $0 \leq x_1 \leq 1, \ldots, 0 \leq x_K \leq 1$ and $x_1 + \cdots + x_K \leq 1$. The corresponding KLD is

$$\begin{aligned} KLD = {} & \log\left[\frac{C(a_1, \ldots, a_K, b, c)}{C(d_1, \ldots, d_K, e, f)}\right] + \sum_{i=1}^{K} (a_i - d_i) \frac{\partial}{\partial \alpha}\left[\frac{C(a_1, \ldots, a_i, \ldots, a_K, b, c)}{C(a_1, \ldots, a_i + \alpha, \ldots, a_K, b, c)}\right]\Bigg|_{\alpha=0} \\ & + (b - e) \frac{\partial}{\partial \alpha}\left[\frac{C(a_1, \ldots, a_i, \ldots, a_K, b, c)}{C(a_1, \ldots, a_K, b + \alpha, c)}\right]\Bigg|_{\alpha=0} \\ & - (c - f) \frac{\partial}{\partial \alpha}\left[\frac{C(a_1, \ldots, a_i, \ldots, a_K, b, c)}{C(a_1, \ldots, a_K, b, c - \alpha)}\right]\Bigg|_{\alpha=0}. \end{aligned}$$

2.5. Multivariate Kotz Type Distribution [12]

Consider the joint probability density functions

$$f_1(x_1,\ldots,x_p) = \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\left(\mathbf{x}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{x}\right)^{N-1}\exp\left\{-q\left(\mathbf{x}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{x}\right)^a\right\}$$

and

$$f_2(x_1,\ldots,x_p) = \frac{b\Gamma\left(\frac{p}{2}\right)s^{\frac{2M+p-2}{2b}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2M+p-2}{2b}\right)}|\boldsymbol{\Sigma}_2|^{-\frac{1}{2}}\left(\mathbf{x}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{x}\right)^{M-1}\exp\left\{-s\left(\mathbf{x}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{x}\right)^b\right\}$$

for $-\infty < x_1 < \infty, \ldots, -\infty < x_p < \infty$, $a > 0$, $q > 0$, $N > 1 - \frac{p}{2}$, $b > 0$, $s > 0$, $M > 1 - \frac{p}{2}$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ positive definite symmetric matrices. The corresponding KLD is

$$\begin{aligned}
KLD = {} & \log\left[\frac{aq^{\frac{2N+p-2}{2a}}\Gamma\left(\frac{2M+p-2}{2b}\right)}{bs^{\frac{2M+p-2}{2b}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\right] + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|} + \frac{N-1}{a}\left[\psi\left(\frac{2N+p-2}{2a}\right) - \log q\right] \\
& - \frac{2N+p-2}{2a} - (M-1)\frac{\Gamma\left(\frac{p}{2}\right)}{a\pi^{\frac{p}{2}}}\left\{\psi\left(\frac{p+2N-2}{2a}\right) - \log q\right\}\left[\prod_{j=1}^{p-1} B\left(\frac{p-j}{2},\frac{1}{2}\right)\right] \\
& - (M-1)\frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}}\log\lambda_1\left[\prod_{j=1}^{p-1} B\left(\frac{p-j}{2},\frac{1}{2}\right)\right] \\
& + (M-1)\frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}}\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{i_1+\cdots+i_{p-1}=k}\binom{k}{i_1,\ldots,i_{p-1}} \\
& \quad \times \prod_{j=1}^{p-1}\left\{\left[\frac{(\lambda_{j+1}-\lambda_j)}{\lambda_1}\right]^{i_j} B\left(\sum_{\ell=j}^{p-1} i_\ell + \frac{p-j}{2},\frac{1}{2}\right)\right\} \\
& + \frac{s\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{2N+p+2b-2}{2a}\right)}{\pi^{\frac{p}{2}}q^{\frac{b}{a}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{b}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}\lambda_1^{b-r_1} \\
& \quad \times \prod_{j=1}^{p-1}\left[(\lambda_{j+1}-\lambda_j)^{r_j-r_{j+1}} B\left(r_j + \frac{p-j}{2},\frac{1}{2}\right)\right]
\end{aligned}$$

provided that the infinite sum converges.

2.6. Multivariate Logistic Distribution [6]

Consider the joint probability density functions

$$f_1(x_1,\ldots,x_p) = \frac{p!a_1\cdots a_p\exp(-a_1x_1-\cdots-a_px_p)}{\left[1+\exp(-a_1x_1)+\cdots+\exp(-a_px_p)\right]^{p+1}}$$

and

$$f_2(x_1,\ldots,x_p) = \frac{p!b_1\cdots b_p\exp(-b_1x_1-\cdots-b_px_p)}{\left[1+\exp(-b_1x_1)+\cdots+\exp(-b_px_p)\right]^{p+1}}$$

for $-\infty < x_1 < \infty, \ldots, -\infty < x_p < \infty$, $a_1 > 0, \ldots, a_p > 0$ and $b_1 > 0, \ldots, b_p > 0$. The corresponding KLD is

$$
\begin{aligned}
KLD = & \log\left(\frac{a_1 \cdots a_p}{b_1 \cdots b_p}\right) \\
& - (p+1) \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \sum_{i_1+\cdots+i_p=k} \binom{k}{i_1, \ldots, i_p} \left[\prod_{j=1}^{p} \Gamma\left(\frac{a_j + i_j b_j}{a_j}\right)\right] \Gamma\left(1 - \sum_{j=1}^{p} \frac{i_j b_j}{a_j}\right) \\
& - (p+1) \frac{\Gamma^{'}(p+1) - \Gamma(p+1)\Gamma^{'}(1)}{p!}
\end{aligned}
$$

provided that $\sum_{j=1}^{p} \frac{i_j b_j}{a_j} < 1$ and the infinite series converges.

2.7. Multivariate Logistic Distribution [7]

Consider the joint probability density functions

$$
f_1(x_1, \ldots, x_p) = \frac{(b)_p a_1 \cdots a_p \exp(-a_1 x_1 - \cdots - a_p x_p)}{\left[1 + \exp(-a_1 x_1) + \cdots + \exp(-a_p x_p)\right]^{b+p}}
$$

and

$$
f_2(x_1, \ldots, x_p) = \frac{(d)_p c_1 \cdots c_p \exp(-c_1 x_1 - \cdots - c_p x_p)}{\left[1 + \exp(-c_1 x_1) + \cdots + \exp(-c_p x_p)\right]^{d+p}}
$$

for $-\infty < x_1 < \infty, \ldots, -\infty < x_p < \infty$, $a_1 > 0, \ldots, a_p > 0$, $c_1 > 0, \ldots, c_p > 0$, $b > 0$ and $d > 0$. The corresponding KLD is

$$
\begin{aligned}
KLD = & \log\left[\frac{(b)_p a_1 \cdots a_p}{(d)_p b_1 \cdots b_p}\right] \\
& - \frac{d+p}{\Gamma(b)} \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \sum_{i_1+\cdots+i_p=k} \binom{k}{i_1, \ldots, i_p} \left[\prod_{j=1}^{p} \Gamma\left(\frac{a_j + i_j c_j}{a_j}\right)\right] \Gamma\left(b - \sum_{j=1}^{p} \frac{i_j c_j}{a_j}\right) \\
& - (b+p) \frac{\Gamma(b)\Gamma^{'}(b+p) - \Gamma(b+p)\Gamma^{'}(b)}{\Gamma(b)\Gamma(b+p)}
\end{aligned}
$$

provided that $\sum_{j=1}^{p} \frac{i_j c_j}{a_j} < b$ and the infinite series converges.

2.8. Sarabia [8]'s Multivariate Normal Distribution

Consider the joint probability density functions

$$
f_1(x_1, \ldots, x_p) = \frac{\sqrt{a_1 \cdots a_p} \beta_p(c, a_1, \ldots, a_p)}{(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{p}\left(a_i x_i^2\right) + c \prod_{i=1}^{p}\left(a_i x_i^2\right)\right]\right\}
$$

and

$$
f_2(x_1, \ldots, x_p) = \frac{\sqrt{b_1 \cdots b_p} \beta_p(d, b_1, \ldots, b_p)}{(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{p}\left(b_i x_i^2\right) + d \prod_{i=1}^{p}\left(b_i x_i^2\right)\right]\right\}
$$

for $-\infty < x_1 < \infty, \ldots, -\infty < x_p < \infty$, $a_1 > 0, \ldots, a_p > 0$, $b_1 > 0, \ldots, b_p > 0$, $c > 0$ and $d > 0$, where $\beta_p(c, a_1, \ldots, a_p)$ and $\beta_p(d, b_1, \ldots, b_p)$ denote normalizing constants. The corresponding KLD is

$$KLD = \log\left[\frac{\sqrt{a_1\cdots a_p}\beta_p(c,a_1,\ldots,a_p)}{\sqrt{b_1\cdots b_p}\beta_p(d,b_1,\ldots,b_p)}\right] + \left[\frac{1}{2} - \frac{c\beta_p'(c,a_1,\ldots,a_p)}{\beta_p(c,a_1,\ldots,a_p)}\right]\sum_{i=1}^{p}\left(\frac{b_i}{a_i}-1\right) + \left(\frac{db_1\cdots b_p}{a_1\cdots a_p} - c\right)\frac{\beta_p'(c,a_1,\ldots,a_p)}{\beta_p(c,a_1,\ldots,a_p)},$$

where $\beta_p'(c,a_1,\ldots,a_p) = \frac{\partial}{\partial c}\beta_p(c,a_1,\ldots,a_p)$.

2.9. Multivariate Pearson Type II Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{\pi^{-\frac{K}{2}}\Gamma\left(\frac{K}{2}\right)\Gamma\left(\frac{K}{2}+a+b-1\right)}{\Gamma(a)\Gamma\left(\frac{K}{2}+b-1\right)}\left(\sum_{i=1}^{K}x_i^2\right)^{b-1}\left(1-\sum_{i=1}^{K}x_i^2\right)^{a-1}$$

and

$$f_2(\mathbf{x}) = \frac{\pi^{-\frac{K}{2}}\Gamma\left(\frac{K}{2}\right)\Gamma\left(\frac{K}{2}+c+d-1\right)}{\Gamma(c)\Gamma\left(\frac{K}{2}+d-1\right)}\left(\sum_{i=1}^{K}x_i^2\right)^{d-1}\left(1-\sum_{i=1}^{K}x_i^2\right)^{c-1}$$

for $K \geq 2$, $a > 0$, $b > 0$, $c > 0$, $d > 0$ and $0 < x_1^2 + \cdots + x_K^2 < 1$. The corresponding KLD is

$$KLD = \log\left[\frac{\Gamma\left(\frac{K}{2}+a+b-1\right)\Gamma(c)\Gamma\left(\frac{K}{2}+d-1\right)}{\Gamma(a)\Gamma\left(\frac{K}{2}+b-1\right)\Gamma\left(\frac{K}{2}+c+d-1\right)}\right] + (b-d)\left[\psi\left(\frac{K}{2}+b-1\right) - \psi\left(\frac{K}{2}+a+b-1\right)\right] + (a-c)\left[\psi(a) - \psi\left(\frac{K}{2}+a+b-1\right)\right].$$

2.10. Multivariate Selberg Beta Distribution [13]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = C(a,b,c)\left|\prod_{1\leq i<j\leq p}(x_i - x_j)\right|^{2c}\prod_{i=1}^{p}\left[x_i^{a-1}(1-x_i)^{b-1}\right]$$

and

$$f_2(\mathbf{x}) = C(d,e,f)\left|\prod_{1\leq i<j\leq p}(x_i - x_j)\right|^{2f}\prod_{i=1}^{p}\left[x_i^{d-1}(1-x_i)^{e-1}\right]$$

for $a > 0$, $b > 0$, $c > 0$, $d > 0$, $e > 0$, $f > 0$ and $0 < x_1 < 1, \ldots, 0 < x_p < 1$. The corresponding KLD is

$$KLD = \log\left[\frac{C(d,e,f)}{C(a,b,c)}\right] + (a-d)\frac{\partial}{\partial\alpha}\left[\frac{C(a,b,c)}{C(a+\alpha,b,c)}\right]\Bigg|_{\alpha=0} + (b-e)\frac{\partial}{\partial\alpha}\left[\frac{C(a,b,c)}{C(a,b+\alpha,c)}\right]\Bigg|_{\alpha=0} + 2(c-f)\frac{\partial}{\partial\alpha}\left[\frac{C(a,b,c)}{C(a,b,c+\alpha)}\right]\Bigg|_{\alpha=0}.$$

2.11. Multivariate Weighted Exponential Distribution [14]

Consider the joint probability density functions

$$f_1(x_1,\ldots,x_p) = \frac{\left(\sum_{i=1}^p a_i\right)\left(\prod_{i=1}^p a_i\right)}{a_{p+1}} \left\{1 - \exp\left[-a_{p+1}\min(x_1,\ldots,x_p)\right]\right\} \left[\prod_{i=1}^p \exp(-a_i x_i)\right]$$

and

$$f_2(x_1,\ldots,x_p) = \frac{\left(\sum_{i=1}^p b_i\right)\left(\prod_{i=1}^p b_i\right)}{b_{p+1}} \left\{1 - \exp\left[-b_{p+1}\min(x_1,\ldots,x_p)\right]\right\} \left[\prod_{i=1}^p \exp(-b_i x_i)\right]$$

for $x_1 > 0, \ldots, x_p > 0$, $a_1 > 0, \ldots, a_{p+1} > 0$ and $b_1 > 0, \ldots, b_{p+1} > 0$. The corresponding KLD is

$$\begin{aligned} KLD = {} & \log\left[\frac{b_{p+1}\left(\sum_{i=1}^p a_i\right)\left(\prod_{i=1}^p a_i\right)}{a_{p+1}\left(\sum_{i=1}^p b_i\right)\left(\prod_{i=1}^p b_i\right)}\right] + \sum_{i=1}^p (b_i - a_i)\left(\frac{1}{a_i} + \frac{1}{a_{p+1}}\right) \\ & - \sum_{k=1}^\infty \frac{(a_1 + \cdots + a_p)(a_1 + \cdots + a_p + a_{p+1})}{k(a_1 + \cdots + a_p + k a_{p+1})\left[a_1 + \cdots + a_p + (k+1)a_{p+1}\right]} \\ & + \sum_{k=1}^\infty \frac{(a_1 + \cdots + a_p)(a_1 + \cdots + a_p + a_{p+1})}{k(a_1 + \cdots + a_p + k b_{p+1})\left[a_1 + \cdots + a_p + a_{p+1} + k b_{p+1}\right]}, \end{aligned}$$

which follows from properties stated in [14] provided that the infinite series converge.

2.12. Von Mises Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{\kappa_1^{\frac{p}{2}-1} \exp\left(\kappa_1 \boldsymbol{\mu}_1^T \mathbf{x}\right)}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa_1)}$$

and

$$f_2(\mathbf{x}) = \frac{\kappa_2^{\frac{p}{2}-1} \exp\left(\kappa_2 \boldsymbol{\mu}_2^T \mathbf{x}\right)}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa_2)}$$

for $\kappa_1 > 0$, $\kappa_2 > 0$, $\boldsymbol{\mu}_1^T \boldsymbol{\mu}_1 = 1$, $\boldsymbol{\mu}_2^T \boldsymbol{\mu}_2 = 1$ and $\mathbf{x}^T \mathbf{x} = 1$. The corresponding KLD is

$$KLD = \log\left[\frac{\kappa_1^{\frac{p}{2}-1}}{\kappa_2^{\frac{p}{2}-1}} \frac{I_{\frac{p}{2}-1}(\kappa_2)}{I_{\frac{p}{2}-1}(\kappa_1)}\right] + \left(\kappa_1 - \kappa_2 \boldsymbol{\mu}_2^T \boldsymbol{\mu}_1\right).$$

3. Exact Expressions for Matrix-Variate Distributions

In this section, we state exact expressions for (1) for matrix-variate beta, matrix-variate Dirichlet, matrix-variate gamma, matrix-variate Gauss hypergeometric, matrix-variate inverse beta, matrix-variate inverse gamma, matrix-variate Kummer beta, matrix-variate Kummer gamma, matrix-variate normal and matrix-variate two-sided power distributions.

3.1. Matrix-Variate Beta Distribution [15]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{\Omega} - \mathbf{x}|^{a-\frac{p+1}{2}} |\mathbf{x}|^{b-\frac{p+1}{2}}}{|\mathbf{\Omega}|^{a+b} B_p(a,b)}$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{\Omega} - \mathbf{x}|^{c-\frac{p+1}{2}} |\mathbf{x}|^{d-\frac{p+1}{2}}}{|\mathbf{\Omega}|^{c+d} B_p(c,d)}$$

for $a > \frac{p-1}{2}$, $b > \frac{p-1}{2}$, $c > \frac{p-1}{2}$, $d > \frac{p-1}{2}$ and $\mathbf{\Omega}$, $\mathbf{x}$, $\mathbf{\Omega} - \mathbf{x}$ being $p \times p$ positive definite matrices. The corresponding KLD is

$$KLD = \log\left[\frac{|\mathbf{\Omega}|^{c+d} B_p(c,d)}{|\mathbf{\Omega}|^{a+b} B_p(a,b)}\right] + (a-c)\frac{\partial}{\partial\alpha}\left\{\frac{|\mathbf{\Omega}|^{\alpha} B_p(\alpha+a,b)}{B_p(a,b)}\right\}\Bigg|_{\alpha=0} + (b-d)\frac{\partial}{\partial\alpha}\left\{\frac{|\mathbf{\Omega}|^{\alpha} B_p(a,\alpha+b)}{B_p(a,b)}\right\}\Bigg|_{\alpha=0}.$$

3.2. Matrix-Variate Dirichlet Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{B_p(a_1, \ldots, a_n; a_{n+1})} |\mathbf{x}_1|^{a_1-p} \cdots |\mathbf{x}_n|^{a_n-p} \left|\mathbf{I}_p - \sum_{i=1}^{n} \mathbf{x}_i\right|^{a_{n+1}-p}$$

and

$$f_2(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{B_p(b_1, \ldots, b_n; b_{n+1})} |\mathbf{x}_1|^{b_1-p} \cdots |\mathbf{x}_n|^{b_n-p} \left|\mathbf{I}_p - \sum_{i=1}^{n} \mathbf{x}_i\right|^{b_{n+1}-p}$$

for $a_1 > \frac{p-1}{2}, \ldots, a_{n+1} > \frac{p-1}{2}$, $b_1 > \frac{p-1}{2}, \ldots, b_{n+1} > \frac{p-1}{2}$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{I}_p - \mathbf{x}_1 - \cdots - \mathbf{x}_n$ being $p \times p$ positive definite matrices. The corresponding KLD is

$$KLD = \log\left[\frac{B_p(b_1, \ldots, b_n; b_{n+1})}{B_p(a_1, \ldots, a_n; a_{n+1})}\right] + (a_{n+1} - b_{n+1})\frac{\partial}{\partial\alpha}\frac{B_p(a_1, \ldots, a_n; a_{n+1}+\alpha)}{B_p(a_1, \ldots, a_n; a_{n+1})}\Bigg|_{\alpha=0} + \sum_{i=1}^{n}(a_i - b_i)\frac{\partial}{\partial\alpha}\frac{B_p(a_1, \ldots, a_i+\alpha, \ldots, a_n; a_{n+1})}{B_p(a_1, \ldots, a_i, \ldots, a_n; a_{n+1})}\Bigg|_{\alpha=0}.$$

3.3. Matrix-Variate Gamma Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{\Sigma}_1|^{-a}}{b^{ap}\Gamma_p(a)} |\mathbf{x}|^{a-\frac{p+1}{2}} \exp\left[\mathrm{tr}\left(-\frac{1}{b}\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)\right]$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{\Sigma}_2|^{-c}}{d^{cp}\Gamma_p(c)} |\mathbf{x}|^{c-\frac{p+1}{2}} \exp\left[\mathrm{tr}\left(-\frac{1}{d}\mathbf{\Sigma}_2^{-1}\mathbf{x}\right)\right]$$

for $a > \frac{p-1}{2}$, $b > 0$, $c > \frac{p-1}{2}$, $d > 0$ and $\mathbf{x}$, $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ being $p \times p$ positive definite symmetric matrices. The corresponding KLD is

$$KLD = \log\left[\frac{d^{pc}\Gamma_p(c)}{b^{pa}\Gamma_p(a)}\frac{|\mathbf{\Sigma}_2|^{c}}{|\mathbf{\Sigma}_1|^{a}}\right] + \frac{a-c}{\Gamma_p(a)}\frac{\partial}{\partial\alpha}\left[|\mathbf{\Sigma}_1|^{\alpha} b^{p\alpha}\Gamma_p(a+\alpha)\right]\Bigg|_{\alpha=0} + \frac{2ap}{b} - \mathrm{tr}\left[-\frac{2a}{d}\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\right].$$

3.4. Matrix-Variate Gauss Hypergeometric Distribution [16]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{x}|^{a-\frac{p+1}{2}} \left|\mathbf{I}_p - \mathbf{x}\right|^{b-\frac{p+1}{2}}}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B}) \left|\mathbf{I}_p + \mathbf{B}\mathbf{x}\right|^{c}}$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{x}|^{d-\frac{p+1}{2}} \left|\mathbf{I}_p - \mathbf{x}\right|^{e-\frac{p+1}{2}}}{B_p(d,e)\ {}_2F_1(d,f;d+e;-\mathbf{B}) \left|\mathbf{I}_p + \mathbf{B}\mathbf{x}\right|^{f}}$$

for $a > \frac{p-1}{2}$, $b > \frac{p-1}{2}$, $0 \le c < \infty$, $d > \frac{p-1}{2}$, $e > \frac{p-1}{2}$, $0 \le f < \infty$ and $\mathbf{x}$, $\mathbf{I}_p - \mathbf{x}$, $\mathbf{B}$, $\mathbf{I}_p + \mathbf{B}$ being $p \times p$ positive definite matrices, where $\Gamma_p(c)$ and $\Gamma_p(f)$ are assumed to exist. The corresponding KLD is

$$\begin{aligned} KLD = {} & \log\left[\frac{B_p(d,e)\ {}_2F_1(d,f;d+e;-\mathbf{B})}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B})}\right] \\ & + (a-d)\frac{\partial}{\partial\alpha}\frac{B_p(a+\alpha,b)\ {}_2F_1(a+\alpha,c;a+b+\alpha;-\mathbf{B})}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B})}\bigg|_{\alpha=0} \\ & + (b-e)\frac{\partial}{\partial\alpha}\frac{B_p(a,b+\alpha)\ {}_2F_1(a,c;a+b+\alpha;-\mathbf{B})}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B})}\bigg|_{\alpha=0} \\ & + (f-c)\frac{\partial}{\partial\alpha}\frac{{}_2F_1(a,c-\alpha;a+b;-\mathbf{B})}{{}_2F_1(a,c;a+b;-\mathbf{B})}\bigg|_{\alpha=0}. \end{aligned}$$

3.5. Matrix-Variate Inverse Beta Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{\Omega}+\mathbf{x}|^{-a-b}|\mathbf{x}|^{b-\frac{p+1}{2}}}{|\mathbf{\Omega}|^{-a}B_p(a,b)}$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{\Omega}+\mathbf{x}|^{-c-d}|\mathbf{x}|^{d-\frac{p+1}{2}}}{|\mathbf{\Omega}|^{-c}B_p(c,d)}$$

for $a > \frac{p-1}{2}$, $b > \frac{p-1}{2}$, $c > \frac{p-1}{2}$, $d > \frac{p-1}{2}$ and $\mathbf{x}$, $\mathbf{\Omega}$, $\mathbf{\Omega}+\mathbf{x}$ being $p \times p$ positive definite matrices. The corresponding KLD is

$$\begin{aligned} KLD = {} & \log\left[\frac{|\mathbf{\Omega}|^{a-c}B_p(c,d)}{B_p(a,b)}\right] + (c+d-a-b)\frac{\partial}{\partial\alpha}\left\{\frac{|\mathbf{\Omega}|^{\alpha}B_p(a-\alpha,b)}{B_p(a,b)}\right\}\bigg|_{\alpha=0} \\ & + (b-d)\frac{\partial}{\partial\alpha}\left\{\frac{|\mathbf{\Omega}|^{\alpha}B_p(a-\alpha,\alpha+b)}{B_p(a,b)}\right\}\bigg|_{\alpha=0}. \end{aligned}$$

3.6. Matrix-Variate Inverse Gamma Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{\Sigma}_1|^{a}}{b^{ap}\Gamma_p(a)}|\mathbf{x}|^{-a-\frac{p+1}{2}}\exp\left[\mathrm{tr}\left(-\frac{1}{b}\mathbf{\Sigma}_1\mathbf{x}^{-1}\right)\right]$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{\Sigma}_2|^c}{d^{cp}\Gamma_p(c)} |\mathbf{x}|^{-c-\frac{p+1}{2}} \exp\left[\text{tr}\left(-\frac{1}{d}\mathbf{\Sigma}_2\mathbf{x}^{-1}\right)\right]$$

for $a > \frac{p-1}{2}$, $b > 0$, $c > \frac{p-1}{2}$, $d > 0$ and $\mathbf{x}$, $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ being $p \times p$ positive definite symmetric matrices. The corresponding KLD is

$$KLD = \log\left[\frac{d^{pc}\Gamma_p(c)}{b^{pa}\Gamma_p(a)} \frac{|\mathbf{\Sigma}_1|^a}{|\mathbf{\Sigma}_2|^c}\right] + \frac{c-a}{\Gamma_p(a)} \frac{\partial}{\partial\alpha}\left[\frac{|\mathbf{\Sigma}_1|^\alpha \Gamma_p(a-\alpha)}{b^{p\alpha}}\right]\Bigg|_{\alpha=0} + \frac{2a}{d}\,\text{tr}\left[\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}\right] - \frac{2ap}{b}.$$

3.7. Matrix-Variate Kummer Beta Distribution [17]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{x}|^{a-\frac{p+1}{2}} \left|\mathbf{I}_p - \mathbf{x}\right|^{b-\frac{p+1}{2}} \exp[-\text{tr}\,(\mathbf{B}_1\mathbf{x})]}{B_p(a,b)\,{}_1F_1(a, a+b; -\mathbf{B}_1)}$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{x}|^{c-\frac{p+1}{2}} \left|\mathbf{I}_p - \mathbf{x}\right|^{d-\frac{p+1}{2}} \exp[-\text{tr}\,(\mathbf{B}_2\mathbf{x})]}{B_p(c,d)\,{}_1F_1(c; c+d; -\mathbf{B}_2)}$$

for $a > \frac{p-1}{2}$, $b > \frac{p-1}{2}$, $c > \frac{p-1}{2}$, $d > \frac{p-1}{2}$ and $\mathbf{x}$, $\mathbf{I}_p - \mathbf{x}$, $\mathbf{B}_1$, $\mathbf{B}_2$ being $p \times p$ positive definite matrices. The corresponding KLD is

$$\begin{aligned} KLD = \log\left[\frac{B_p(c,d)\,{}_1F_1(c;c+d;-\mathbf{B}_2)}{B_p(a,b)\,{}_1F_1(a,a+b;-\mathbf{B}_1)}\right] &+ (a-c)\frac{\partial}{\partial\alpha}\frac{B_p(a+\alpha,b)\,{}_1F_1(a+\alpha;a+b+\alpha;-\mathbf{B}_1)}{B_p(a,b)\,{}_1F_1(a;a+b;-\mathbf{B}_1)}\Bigg|_{\alpha=0} \\ &+ (b-d)\frac{\partial}{\partial\alpha}\frac{B_p(a,b+\alpha)\,{}_1F_1(a;a+b+\alpha;-\mathbf{B}_1)}{B_p(a,b)\,{}_1F_1(a;a+b;-\mathbf{B}_1)}\Bigg|_{\alpha=0} \\ &+ \text{tr}\left[(\mathbf{B}_2-\mathbf{B}_1)\frac{\partial}{\partial\mathbf{z}}\frac{{}_1F_1(a;a+b;\mathbf{z}-\mathbf{B}_1)}{{}_1F_1(a;a+b;-\mathbf{B}_1)}\Bigg|_{\mathbf{z}=\mathbf{0}}\right]. \end{aligned}$$

3.8. Matrix-Variate Kummer Gamma Distribution [18]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{|\mathbf{x}|^{a-\frac{p+1}{2}} \left|\mathbf{I}_p + \mathbf{x}\right|^{-b} \exp[-\text{tr}\,(\mathbf{B}_1\mathbf{x})]}{\Gamma_p(a)\,\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}$$

and

$$f_2(\mathbf{x}) = \frac{|\mathbf{x}|^{c-\frac{p+1}{2}} \left|\mathbf{I}_p + \mathbf{x}\right|^{-d} \exp[-\text{tr}\,(\mathbf{B}_2\mathbf{x})]}{\Gamma_p(c)\,\Psi_1\left(c; c-d+\frac{p+1}{2}; \mathbf{B}_2\right)}$$

for $a > \frac{p-1}{2}$, $-\infty < b < \infty$, $c > \frac{p-1}{2}$, $-\infty < d < \infty$ and $\mathbf{x}$, $\mathbf{I}_p + \mathbf{x}$, $\mathbf{B}_1$, $\mathbf{B}_2$ being $p \times p$ positive definite matrices, where $\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)$ and $\Psi_1\left(c; c-d+\frac{p+1}{2}; \mathbf{B}_2\right)$ denote Kummer functions with matrix arguments and the parameters are chosen such that these functions exist. The corresponding KLD is

$$KLD = \log\left[\frac{\Gamma_p(c)\ \Psi_1\left(c; c-d+\frac{p+1}{2}; \mathbf{B}_2\right)}{\Gamma_p(a)\ \Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\right]$$
$$+ (a-c)\,\frac{\partial}{\partial\alpha}\,\frac{\Gamma_p(a+\alpha)\ \Psi_1\left(a+\alpha; a-b+\alpha+\frac{p+1}{2}; \mathbf{B}_1\right)}{\Gamma_p(a)\ \Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\Bigg|_{\alpha=0}$$
$$+ (d-p)\,\frac{\partial}{\partial\alpha}\,\frac{\Psi_1\left(a; a-b+\alpha+\frac{p+1}{2}; \mathbf{B}_1\right)}{\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\Bigg|_{\alpha=0}$$
$$+ \operatorname{tr}\left[(\mathbf{B}_2-\mathbf{B}_1)\,\frac{\partial}{\partial\mathbf{z}}\,\frac{\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1-\mathbf{z}\right)}{\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\Bigg|_{\mathbf{z}=\mathbf{0}}\right].$$

3.9. Matrix-Variate Normal Distribution

Consider the joint probability density functions

$$f_1(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\mathbf{V}_1|^{\frac{n}{2}}|\mathbf{U}_1|^{\frac{p}{2}}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left[\mathbf{V}_1^{-1}(\mathbf{x}-\mathbf{M}_1)^T\mathbf{U}_1^{-1}(\mathbf{x}-\mathbf{M}_1)\right]\right\}$$

and

$$f_2(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\mathbf{V}_2|^{\frac{n}{2}}|\mathbf{U}_2|^{\frac{p}{2}}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left[\mathbf{V}_2^{-1}(\mathbf{x}-\mathbf{M}_2)^T\mathbf{U}_2^{-1}(\mathbf{x}-\mathbf{M}_2)\right]\right\}$$

for $\mathbf{U}_1$, $\mathbf{U}_2$ being positive definite symmetric matrices of dimension $n \times n$, $\mathbf{V}_1$, $\mathbf{V}_2$ being positive definite symmetric matrices of dimension $p \times p$ and $\mathbf{M}_1$, $\mathbf{M}_2$ being matrices of dimension $n \times p$. The corresponding KLD is

$$KLD = \log\left[\frac{|\mathbf{V}_2|^{\frac{n}{2}}|\mathbf{U}_2|^{\frac{p}{2}}}{|\mathbf{V}_1|^{\frac{n}{2}}|\mathbf{U}_1|^{\frac{p}{2}}}\right] + \frac{1}{2}\operatorname{tr}(\mathbf{U}_1)\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{V}_1\mathbf{U}_2^{-1}\right) + \frac{1}{2}\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_1^T\mathbf{M}_1\mathbf{U}_2^{-1}\right)$$
$$- \frac{1}{2}\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_1^T\mathbf{M}_2\mathbf{U}_2^{-1}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_2^T\mathbf{M}_1\mathbf{U}_2^{-1}\right)$$
$$+ \frac{1}{2}\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_2^T\mathbf{M}_2\mathbf{U}_2^{-1}\right) - \frac{1}{2}\operatorname{tr}(\mathbf{U}_1)\operatorname{tr}\left(\mathbf{U}_1^{-1}\right).$$

3.10. Matrix-Variate Two-Sided Power Distribution [19]

Consider the joint probability density functions

$$f_1(\mathbf{x}) = C(a)\begin{cases} |\mathbf{x}|^{a-\frac{p+1}{2}}|\mathbf{B}|^{-a+\frac{p+1}{2}}, & \mathbf{0}_p < \mathbf{x} \le \mathbf{B}, \\ \left|\mathbf{I}_p-\mathbf{x}\right|^{a-\frac{p+1}{2}}\left|\mathbf{I}_p-\mathbf{B}\right|^{-a+\frac{p+1}{2}}, & \mathbf{B} \le \mathbf{x} \le \mathbf{I}_p \end{cases}$$

and

$$f_2(\mathbf{x}) = C(b)\begin{cases} |\mathbf{x}|^{b-\frac{p+1}{2}}|\mathbf{B}|^{-b+\frac{p+1}{2}}, & \mathbf{0}_p < \mathbf{x} \le \mathbf{B}, \\ \left|\mathbf{I}_p-\mathbf{x}\right|^{b-\frac{p+1}{2}}\left|\mathbf{I}_p-\mathbf{B}\right|^{-b+\frac{p+1}{2}}, & \mathbf{B} \le \mathbf{x} \le \mathbf{I}_p, \end{cases}$$

for $a > \frac{p-1}{2}$, $b > \frac{p-1}{2}$ and $\mathbf{x}$, $\mathbf{I}_p - \mathbf{x}$, $\mathbf{B}$, $\mathbf{I}_p - \mathbf{B}$ being $p \times p$ positive definite matrices, where

$$C(a) = |\mathbf{B}|^{\frac{p+1}{2}}B_p\left(a, \frac{p+1}{2}\right) + \left|\mathbf{I}_p-\mathbf{B}\right|^{\frac{p+1}{2}}B_p\left(\frac{p+1}{2}, a\right)$$

and

$$C(b) = |\mathbf{B}|^{\frac{p+1}{2}} B_p\left(b, \frac{p+1}{2}\right) + \left|\mathbf{I}_p - \mathbf{B}\right|^{\frac{p+1}{2}} B_p\left(\frac{p+1}{2}, b\right).$$

The corresponding KLD is

$$\begin{aligned} KLD = \log\left[\frac{C(a)}{C(b)}\right] &+ (a-b)\frac{\partial}{\partial\alpha}\left\{|\mathbf{B}|^{\alpha+\frac{p+1}{2}} B_p\left(a+\alpha, \frac{p+1}{2}\right)\right\}\Big|_{\alpha=0} \\ &+ (a-b)\frac{\partial}{\partial\alpha}\left\{\left|\mathbf{I}_p - \mathbf{B}\right|^{\alpha+\frac{p+1}{2}} B_p\left(\frac{p+1}{2}, a+\alpha\right)\right\}\Big|_{\alpha=0} \\ &+ (b-a)|\mathbf{B}|^{\frac{p+1}{2}} \log|\mathbf{B}| B_p\left(a, \frac{p+1}{2}\right) \\ &+ (b-a)\left|\mathbf{I}_p - \mathbf{B}\right|^{\frac{p+1}{2}} \log\left|\mathbf{I}_p - \mathbf{B}\right| B_p\left(\frac{p+1}{2}, a\right). \end{aligned}$$

4. Proofs

Before presenting the proofs of the expressions in Sections 2 and 3, we state a lemma and give its proof.

4.1. A Technical Lemma

Lemma 1. *Let*

$$I\left(a_1, \ldots, a_p, t_1, \ldots, t_p, b\right) = \int_{\mathbb{R}^p} \frac{\exp\left(-\sum_{j=1}^{p} t_j x_j\right)}{\left[1 + \sum_{j=1}^{p} \exp\left(-a_j x_j\right)\right]^b} dx_1 \cdots dx_p$$

for $t_j > 0$, $a_j > 0$, $j = 1, 2, \ldots, p$ and $b > 0$. Then,

$$I\left(a_1, \ldots, a_p, t_1, \ldots, t_p, b\right) = \frac{1}{a_1 \cdots a_p \Gamma(b)}\left[\prod_{j=1}^{p} \Gamma\left(\frac{t_j}{a_j}\right)\right] \Gamma\left(b - \sum_{j=1}^{p} \frac{t_j}{a_j}\right)$$

provided that $b - \sum_{j=1}^{p} \frac{t_j}{a_j} > 0$.

Proof. Setting $y_j = \exp\left(-a_j x_j\right)$ and assuming the conditions in the lemma, we can write

$$\begin{aligned} &I\left(a_1, \ldots, a_p, t_1, \ldots, t_p, b\right) \\ &= \frac{1}{\Gamma(b)} \int_{\mathbb{R}^p} \int_0^\infty t^{b-1} \exp\left\{-\left[1 + \sum_{j=1}^{p} \exp\left(-a_j x_j\right)\right] t\right\} \exp\left(-\sum_{j=1}^{p} t_j x_j\right) dt dx_1 \cdots dx_p \\ &= \frac{1}{\Gamma(b)} \int_0^\infty t^{b-1} e^{-t} \prod_{j=1}^{p} \left\{\int_{-\infty}^{\infty} \exp\left[-t_j x_j - t\exp\left(-a_j x_j\right)\right] dx_j\right\} dt \\ &= \frac{1}{a_1 \cdots a_p \Gamma(b)} \int_0^\infty t^{b-1} e^{-t} \prod_{j=1}^{p} \left[\int_0^\infty y_j^{\frac{t_j}{a_j}-1} \exp\left(-t y_j\right) dy_j\right] dt \\ &= \frac{1}{a_1 \cdots a_p \Gamma(b)} \left[\prod_{j=1}^{p} \Gamma\left(\frac{t_j}{a_j}\right)\right] \int_0^\infty t^{b-\sum_{j=1}^{p} \frac{t_j}{a_j}-1} e^{-t} dt. \end{aligned}$$

The result follows. □

4.2. Proof for Section 2.1

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{B(b_1,\ldots,b_{K-1};b_K)}{B(a_1,\ldots,a_{K-1};a_K)}\right] + \sum_{i=1}^{K}(a_i - b_i)E[\log X_i]. \quad (2)$$

It is easy to show that

$$E[\log X_i] = \psi(a_i) - \psi(a_1 + \cdots + a_K),$$

so (2) reduces to the required.

4.3. Proof for Section 2.2

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{\alpha\Gamma\left(\frac{p}{\beta}\right)}{\beta\Gamma\left(\frac{p}{\alpha}\right)}\right] + \frac{1}{2}\log\frac{|\mathbf{V}_2|}{|\mathbf{V}_1|} + E\left[\left(\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X}\right)^{\frac{\beta}{2}}\right] - E\left[\left(\mathbf{X}^T\mathbf{V}_1^{-1}\mathbf{X}\right)^{\frac{\alpha}{2}}\right]. \quad (3)$$

The second expectation in (3) can be expressed as

$$\begin{aligned}
E\left[\left(\mathbf{X}^T\mathbf{V}_1^{-1}\mathbf{X}\right)^{\frac{\alpha}{2}}\right] &= \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}|\mathbf{V}_1|^{-\frac{1}{2}}\int_{\mathbb{R}^p}\left(\mathbf{x}^T\mathbf{V}_1^{-1}\mathbf{x}\right)^{\frac{\alpha}{2}}\exp\left\{-\left(\mathbf{x}^T\mathbf{V}_1^{-1}\mathbf{x}\right)^{\frac{\alpha}{2}}\right\}d\mathbf{x} \\
&= \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_{\mathbb{R}^p}\left(\mathbf{y}^T\mathbf{y}\right)^{\frac{\alpha}{2}}\exp\left\{-\left(\mathbf{y}^T\mathbf{y}\right)^{\frac{\alpha}{2}}\right\}d\mathbf{y} \\
&= \frac{\alpha}{2\Gamma\left(\frac{p}{\alpha}\right)}\int_0^\infty u^{\frac{p}{2}+\frac{\alpha}{2}-1}\exp\left(-u^{\frac{\alpha}{2}}\right)du \\
&= \frac{1}{\Gamma\left(\frac{p}{\alpha}\right)}\int_0^\infty t^{\frac{p}{\alpha}}\exp(-t)dt \\
&= \frac{p}{\alpha},
\end{aligned}$$

where $\mathbf{y} = \mathbf{V}_1^{-\frac{1}{2}}\mathbf{x}$, $u = \mathbf{y}^T\mathbf{y}$ and $t = u^{\frac{\alpha}{2}}$.

Let $\mathbf{V} = \mathbf{V}_1^{\frac{1}{2}}\mathbf{V}_2^{-1}\mathbf{V}_1^{\frac{1}{2}}$ and $\mathbf{V} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, where $\mathbf{P}$ is an orthonormal matrix composed of eigenvectors of $\mathbf{V}$ and $\mathbf{D}$ is a diagonal matrix composed of eigenvalues say λ_i of $\mathbf{V}$. Then, the first expectation in (3) can be expressed as

$$\begin{aligned}
E\left[\left(\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X}\right)^{\frac{\beta}{2}}\right] &= \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_{\mathbb{R}^p}\left[\mathrm{tr}\left(\mathbf{D}\mathbf{P}^T\mathbf{y}\mathbf{y}^T\mathbf{P}\right)\right]^{\frac{\beta}{2}}\exp\left\{-\left(\mathbf{y}^T\mathbf{y}\right)^{\frac{\alpha}{2}}\right\}d\mathbf{y} \\
&= \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_{\mathbb{R}^p}\left(\mathbf{x}^T\mathbf{D}\mathbf{x}\right)^{\frac{\beta}{2}}\exp\left\{-\left(\mathbf{x}^T\mathbf{x}\right)^{\frac{\alpha}{2}}\right\}d\mathbf{x} \\
&= \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\left(\sum_i^p\lambda_i x_i^2\right)^{\frac{\beta}{2}}\exp\left\{-\left(\sum_i^p x_i^2\right)^{\frac{\alpha}{2}}\right\}d\mathbf{x}, \quad (4)
\end{aligned}$$

where $\mathbf{y} = \mathbf{V}_1^{-\frac{1}{2}}\mathbf{x}$ and $\mathbf{z} = \mathbf{P}^T\mathbf{y}$. Using the pseudo-polar transformation $z_1 = r\sin\theta_1$, $z_2 = r\cos\theta_1\sin\theta_2, \ldots, z_p = r\cos\theta_1\cos\theta_2\cdots\cos\theta_{p-1}$ https://en.wikipedia.org/wiki/Polar_coordinate_system (accessed on 1 July 2024), (4) can be expressed as

$$E\left[\left(\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X}\right)^{\frac{\beta}{2}}\right] = \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{2\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_0^\infty r^{p-1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\pi}^{\pi}\left[r^2\left(\lambda_1\sin^2\theta_1+\cdots+\lambda_p\cos^2\theta_1\cdots\cos^2\theta_{p-1}\right)\right]^{\frac{\beta}{2}}$$
$$\times\exp\left[-\left(r^2\right)^{\frac{\alpha}{2}}\right]\left[\prod_{j=1}^{p-1}\left|\cos\theta_j\right|^{p-j-1}\right]dr\prod_{j=1}^{p-1}d\theta_j$$
$$= \frac{\alpha\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_0^\infty r^{p+\beta-1}\exp(-r^\alpha)\int_0^1\cdots\int_0^1\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]$$
$$\times\left[\mathcal{B}_p(x_1,\ldots,x_{p-1})\right]^{\frac{\beta}{2}}dx_1\cdots dx_{p-1}dr$$
$$= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{p+\beta}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\}\left[\mathcal{B}_p(x_1,\ldots,x_{p-1})\right]^{\frac{\beta}{2}}dx_1\cdots dx_{p-1}, \tag{5}$$

where $x_i=\cos^2\theta_i$ and $\mathcal{B}_p(x_1,\ldots,x_{p-1})=\lambda_1+(\lambda_2-\lambda_1)x_1+\cdots+\left(\lambda_p-\lambda_{p-1}\right)x_1x_2\cdots x_{p-1}$. Provided that $|\lambda_1|\geq\left|(\lambda_2-\lambda_1)x_1+\cdots+\left(\lambda_p-\lambda_{p-1}\right)x_1x_2\cdots x_{p-1}\right|$ holds, we can apply the generalized multinomial theorem to calculate (5) as

$$E\left[\left(\mathbf{X}^T\mathbf{V}_2^{-1}\mathbf{X}\right)^{\frac{\beta}{2}}\right] = \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{p+\beta}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\}$$
$$\times\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{\frac{\beta}{2}}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}$$
$$\times\lambda_1^{\frac{\beta}{2}-r_1}\left[(\lambda_2-\lambda_1)x_1\right]^{r_1-r_2}\left[(\lambda_3-\lambda_2)x_1x_2\right]^{r_2-r_3}$$
$$\cdots\left[\left(\lambda_{p-1}-\lambda_{p-2}\right)x_1x_2\cdots x_{p-2}\right]^{r_{p-2}-r_{p-1}}\left[\left(\lambda_p-\lambda_{p-1}\right)x_1x_2\cdots x_{p-1}\right]^{r_{p-1}}dx_1\cdots dx_{p-1}$$
$$= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{p+\beta}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\}$$
$$\times\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{\frac{\beta}{2}}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}\lambda_1^{\frac{\beta}{2}-r_1}$$
$$\times\prod_{j=1}^{p-1}\left[\left(\lambda_{j+1}-\lambda_j\right)\left(\prod_{\ell=1}^{j}x_\ell\right)\right]^{r_j-r_{j+1}}dx_1\cdots dx_{p-1}$$
$$= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{p+\beta}{2}\right)}{\pi^{\frac{p}{2}}\Gamma\left(\frac{p}{\alpha}\right)}\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{\frac{\beta}{2}}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}\lambda_1^{\frac{\beta}{2}-r_1}$$
$$\times\prod_{j=1}^{p-1}\left[\left(\lambda_{j+1}-\lambda_j\right)^{r_j-r_{j+1}}B\left(r_j+\frac{p-j}{2},\frac{1}{2}\right)\right]$$

provided that the infinite sum converges. Hence, the required.

4.4. Proof for Section 2.3

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{\Gamma(a_1+\cdots+a_{K+1})\Gamma(b_1)\cdots\Gamma(b_{K+1})}{\Gamma(b_1+\cdots+b_{K+1})\Gamma(a_1)\cdots\Gamma(a_{K+1})}\right]+\sum_{i=1}^{K}(a_i-b_i)E[\log X_i]$$
$$+(b_1+\cdots+b_{K+1}-a_1-\cdots-a_{K+1})E\left[\log\left(1+\sum_{i=1}^{K}X_i\right)\right]. \tag{6}$$

It is easy to show that

$$E[\log X_i] = \psi(a_i) - \psi(a_{K+1})$$

and

$$E\left[\log\left(1 + \sum_{i=1}^{K} X_i\right)\right] = \psi(a_1 + \cdots + a_{K+1}) - \psi(a_{K+1}),$$

so (6) reduces to the required.

4.5. Proof for Section 2.4

The corresponding KLD can be expressed as

$$\begin{aligned} KLD = \log\left[\frac{C(a_1,\ldots,a_K,b,c)}{C(d_1,\ldots,d_K,e,f)}\right] + \sum_{i=1}^{K}(a_i - d_i)E[\log X_i] + (b-e)E\left[\log\left(1 - \sum_{i=1}^{K} X_i\right)\right] \\ - (c-f)E\left[\log\left(1 + \sum_{i=1}^{K} X_i\right)\right]. \end{aligned} \tag{7}$$

It is easy to show that

$$E[\log X_i] = \frac{\partial}{\partial\alpha}\left[\frac{C(a_1,\ldots,a_i,\ldots,a_K,b,c)}{C(a_1,\ldots,a_i+\alpha,\ldots,a_K,b,c)}\right]\Bigg|_{\alpha=0},$$

$$E\left[\log\left(1 - \sum_{i=1}^{K} X_i\right)\right] = \frac{\partial}{\partial\alpha}\left[\frac{C(a_1,\ldots,a_i,\ldots,a_K,b,c)}{C(a_1,\ldots,a_K,b+\alpha,c)}\right]\Bigg|_{\alpha=0}$$

and

$$E\left[\log\left(1 + \sum_{i=1}^{K} X_i\right)\right] = \frac{\partial}{\partial\alpha}\left[\frac{C(a_1,\ldots,a_i,\ldots,a_K,b,c)}{C(a_1,\ldots,a_K,b,c-\alpha)}\right]\Bigg|_{\alpha=0},$$

so (7) reduces to the required.

4.6. Proof for Section 2.5

The corresponding KLD can be expressed as

$$\begin{aligned} KLD = \log\left[\frac{aq^{\frac{2N+p-2}{2a}}\Gamma\left(\frac{2M+p-2}{2b}\right)}{bs^{\frac{2M+p-2}{2b}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\right] + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|} + (N-1)E\left[\log\left(\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\right)\right] \\ - qE\left[\left(\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\right)^a\right] - (M-1)E\left[\log\left(\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X}\right)\right] + sE\left[\left(\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X}\right)^b\right]. \end{aligned} \tag{8}$$

The second expectation in (8) can be calculated as

$$\begin{aligned} E\left[\left(\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\right)^a\right] &= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}|\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}\int_{\mathbb{R}^p}\left(\mathbf{x}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{x}\right)^{a+N-1}\exp\left\{-q\left(\mathbf{x}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{x}\right)^a\right\}d\mathbf{x} \\ &= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\mathbf{y}^T\mathbf{y}\right)^{a+N-1}\exp\left\{-q\left(\mathbf{y}^T\mathbf{y}\right)^a\right\}d\mathbf{y} \\ &= \frac{1}{q\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty t^{\frac{2N+p-2}{2a}}\exp(-t)dt \\ &= \frac{2N+p-2}{2aq}, \end{aligned}$$

where $\mathbf{y} = \mathbf{\Sigma}_1^{-\frac{1}{2}}\mathbf{x}$ and $t = q(\mathbf{y}^T\mathbf{y})^a$.

Let $\mathbf{\Sigma} = \mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1^{\frac{1}{2}}$. The first expectation in (8) can be calculated as

$$\begin{aligned}
E\left[\log\left(\mathbf{X}^T\mathbf{\Sigma}_1^{-1}\mathbf{X}\right)\right] &= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}|\mathbf{\Sigma}_1|^{-\frac{1}{2}}\int_{\mathbb{R}^p}\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)^{N-1}\log\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)\exp\left\{-q\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)^a\right\}d\mathbf{x} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\mathbf{y}^T\mathbf{y}\right)^{N-1}\log\left(\mathbf{y}^T\mathbf{y}\right)\exp\left\{-q\left(\mathbf{y}^T\mathbf{y}\right)^a\right\}d\mathbf{y} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\frac{\partial}{\partial N}\int_{\mathbb{R}^p}\left(\mathbf{y}^T\mathbf{y}\right)^{N-1}\exp\left\{-q\left(\mathbf{y}^T\mathbf{y}\right)^a\right\}d\mathbf{y} \\
&= \frac{q^{\frac{2N+p-2}{2a}}}{\Gamma\left(\frac{2N+p-2}{2a}\right)}\frac{\partial}{\partial N}\int_0^{\infty}\frac{t^{\frac{2N+p-2}{2a}-1}}{q^{\frac{2N+p-2}{2a}}}\exp(-t)dt \\
&= \frac{q^{\frac{2N+p-2}{2a}}}{\Gamma\left(\frac{2N+p-2}{2a}\right)}\frac{\partial}{\partial N}\left[\frac{\Gamma\left(\frac{2N+p-2}{2a}\right)}{q^{\frac{2N+p-2}{2a}}}\right] \\
&= \frac{1}{a}\left[\psi\left(\frac{2N+p-2}{2a}\right)-\log q\right],
\end{aligned}$$

where $\mathbf{y} = \mathbf{\Sigma}_1^{-\frac{1}{2}}\mathbf{x}$ and $t = q(\mathbf{y}^T\mathbf{y})^a$.

As in Section 4.3, write $\mathbf{\Sigma} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, where $\mathbf{P}$ is an orthonormal matrix composed of eigenvectors of $\mathbf{\Sigma}$ and $\mathbf{D}$ is a diagonal matrix composed of eigenvalues say λ_i of $\mathbf{\Sigma}$. Then, the fourth expectation in (8) can be expressed as

$$\begin{aligned}
E\left[\left(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X}\right)^b\right] &= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}|\mathbf{\Sigma}_1|^{-\frac{1}{2}}\int_{\mathbb{R}^p}\left(\mathbf{x}^T\mathbf{\Sigma}_2^{-1}\mathbf{x}\right)^b\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)^{N-1}\exp\left\{-q\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)^a\right\}d\mathbf{x} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\mathbf{y}^T\mathbf{\Sigma}\mathbf{y}\right)^b\left(\mathbf{y}^T\mathbf{y}\right)^{N-1}\exp\left\{-q\left(\mathbf{y}^T\mathbf{y}\right)^a\right\}d\mathbf{y} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left[\operatorname{tr}\left(\mathbf{D}\mathbf{P}^T\mathbf{y}\mathbf{y}^T\mathbf{P}\right)\right]^b\left(\mathbf{y}^T\mathbf{y}\right)^{N-1}\exp\left\{-q\left(\mathbf{y}^T\mathbf{y}\right)^a\right\}d\mathbf{y} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\mathbf{v}^T\mathbf{D}\mathbf{v}\right)^b\left(\mathbf{v}^T\mathbf{v}\right)^{N-1}\exp\left\{-q\left(\mathbf{v}^T\mathbf{v}\right)^a\right\}d\mathbf{v} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\sum_{i=1}^p\lambda_i v_i^2\right)^b\left(\sum_{i=1}^p v_i^2\right)^{N-1}\exp\left\{-q\left(\sum_{i=1}^p v_i^2\right)^a\right\}d\mathbf{v},
\end{aligned} \tag{9}$$

where $\mathbf{y} = \mathbf{\Sigma}_1^{-\frac{1}{2}}\mathbf{x}$ and $\mathbf{v} = \mathbf{P}^T\mathbf{y}$.

Using the pseudo-polar transformation $v_1 = r\sin\theta_1$, $v_2 = r\cos\theta_1\sin\theta_2, \ldots, v_p = r\cos\theta_1\cos\theta_2\cdots\cos\theta_{p-1}$, (9) can be expressed as

$$
\begin{aligned}
E\left[\left(\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X}\right)^b\right] &= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty r^{p-1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\pi}^{\pi}\left[r^2\left(\lambda_1\sin^2\theta_1+\cdots+\lambda_p\cos^2\theta_1\cdots\cos^2\theta_{p-1}\right)\right]^b \\
&\quad\times\left(r^2\right)^{N-1}\exp\left[-q\left(r^2\right)^a\right]\left[\prod_{j=1}^{p-1}\left|\cos\theta_j\right|^{p-j-1}\right]dr\prod_{j=1}^{p-1}d\theta_j \\
&= \frac{\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty\frac{t^{\frac{2N+p+2b-2}{2a}-1}}{q^{\frac{2N+p+2b-2}{2a}}}\exp(-t)\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\} \\
&\quad\times\left[\mathcal{B}_p\left(x_1,\ldots,x_{p-1}\right)\right]^b dx_1\cdots dx_{p-1}dt \\
&= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{2N+p+2b-2}{2a}\right)}{\pi^{\frac{p}{2}}q^{\frac{b}{a}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\} \\
&\quad\times\left[\mathcal{B}_p\left(x_1,\ldots,x_{p-1}\right)\right]^b dx_1\cdots dx_{p-1},
\end{aligned}
\tag{10}
$$

where $x_i=\cos^2\theta_i$, $t=qr^{2a}$ and $\mathcal{B}_p(x_1,\ldots,x_{p-1})=\lambda_1+(\lambda_2-\lambda_1)x_1+\cdots+\left(\lambda_p-\lambda_{p-1}\right)$ $x_1x_2\cdots x_{p-1}$.

Provided that $|\lambda_1|\geq\left|(\lambda_2-\lambda_1)x_1+\cdots+\left(\lambda_p-\lambda_{p-1}\right)x_1x_2\cdots x_{p-1}\right|$ holds, we can apply the generalized multinomial theorem to calculate (10) as

$$
\begin{aligned}
E\left[\left(\mathbf{X}^T\boldsymbol{\Sigma}_2^{-1}\mathbf{X}\right)^b\right] &= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{2N+p+2b-2}{2a}\right)}{\pi^{\frac{p}{2}}q^{\frac{b}{a}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\} \\
&\quad\times\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{b}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}\lambda_1^{b-r_1}\left[(\lambda_2-\lambda_1)x_1\right]^{r_1-r_2}\left[(\lambda_3-\lambda_2)x_1x_2\right]^{r_2-r_3} \\
&\quad\cdots\left[\left(\lambda_{p-1}-\lambda_{p-2}\right)x_1x_2\cdots x_{p-2}\right]^{r_{p-2}-r_{p-1}}\left[\left(\lambda_p-\lambda_{p-1}\right)x_1x_2\cdots x_{p-1}\right]^{r_{p-1}}dx_1\cdots dx_{p-1} \\
&= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{2N+p+2b-2}{2a}\right)}{\pi^{\frac{p}{2}}q^{\frac{b}{a}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}\left(1-x_j\right)^{-\frac{1}{2}}\right]\right\} \\
&\quad\times\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{b}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}} \\
&\quad\times\lambda_1^{b-r_1}\left\{\prod_{j=1}^{p-1}\left[\left(\lambda_{j+1}-\lambda_j\right)\prod_{\ell=1}^{j}x_\ell\right]^{r_j-r_{j+1}}\right\}dx_1\cdots dx_{p-1} \\
&= \frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{2N+p+2b-2}{2a}\right)}{\pi^{\frac{p}{2}}q^{\frac{b}{a}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\sum_{r_1=0}^{\infty}\sum_{r_2=0}^{r_1}\cdots\sum_{r_{p-1}=0}^{r_{p-2}}\binom{b}{r_1}\binom{r_1}{r_2}\cdots\binom{r_{p-2}}{r_{p-1}}\lambda_1^{b-r_1} \\
&\quad\times\prod_{j=1}^{p-1}\left[\left(\lambda_{j+1}-\lambda_j\right)^{r_j-r_{j+1}}B\left(r_j+\frac{p-j}{2},\frac{1}{2}\right)\right]
\end{aligned}
$$

provided that the infinite sum converges.

The third expectation in (8) can be expressed as

$$
\begin{aligned}
E\left[\log\left(\mathbf{X}^T\mathbf{\Sigma}_2^{-1}\mathbf{X}\right)\right] &= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}|\mathbf{\Sigma}_1|^{-\frac{1}{2}}\int_{\mathbb{R}^p}\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)^{N-1}\log\left(\mathbf{x}^T\mathbf{\Sigma}_2^{-1}\mathbf{x}\right)\exp\left\{-q\left(\mathbf{x}^T\mathbf{\Sigma}_1^{-1}\mathbf{x}\right)^a\right\}d\mathbf{x} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\mathbf{y}^T\mathbf{y}\right)^{N-1}\log\left[\mathrm{tr}\left(\mathbf{D}\mathbf{P}^T\mathbf{y}\mathbf{y}^T\mathbf{P}\right)\right]\exp\left\{-q\left(\mathbf{y}^T\mathbf{y}\right)^a\right\}d\mathbf{y} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_{\mathbb{R}^p}\left(\sum_{i=1}^p v_i^2\right)^{N-1}\log\left(\sum_{i=1}^p \lambda_i v_i^2\right)\exp\left\{-q\left(\sum_{i=1}^p v_i^2\right)^a\right\}d\mathbf{v} \\
&= \frac{a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty r^{p-1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cdots\int_{-\pi}^{\pi}\left(r^2\right)^{N-1} \\
&\quad\times\log\left[r^2\left(\lambda_1\sin^2\theta_1+\cdots+\lambda_p\cos^2\theta_1\cdots\cos^2\theta_{p-1}\right)\right] \\
&\quad\times\exp\left\{-q\left(r^2\right)^a\right\}\left[\prod_{j=1}^{p-1}|\cos\theta_j|^{p-j-1}\right]dr\prod_{j=1}^{p-1}d\theta_j \\
&= I_1+I_2
\end{aligned}
$$

say, where $\mathbf{y}=\mathbf{\Sigma}_1^{-\frac{1}{2}}\mathbf{x}$, $\mathbf{v}=\mathbf{P}^T\mathbf{y}$,

$$
\begin{aligned}
I_1 &= \frac{2a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty r^{p+2N-3}\log\left(r^2\right)\exp\left(-qr^{2a}\right) \\
&\quad\times\int_0^1\cdots\int_0^1\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}(1-x_j)^{-\frac{1}{2}}\right]dx_1\cdots dx_{p-1}dr
\end{aligned}
$$

and

$$
\begin{aligned}
I_2 &= \frac{2a\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty r^{p+2N-3}\exp\left(-qr^{2a}\right) \\
&\quad\times\int_0^1\cdots\int_0^1\left\{\prod_{j=1}^{p-1}\left[x_j^{\frac{p-j}{2}-1}(1-x_j)^{-\frac{1}{2}}\right]\right\}\log\left[\mathcal{B}_p(x_1,\ldots,x_{p-1})\right]dx_1\cdots dx_{p-1}dr.
\end{aligned}
$$

The I_1 can be calculated as

$$
\begin{aligned}
I_1 &= \frac{\Gamma\left(\frac{p}{2}\right)q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\int_0^\infty\frac{t^{\frac{p+2N-2}{2a}-1}}{q^{\frac{2N+p-2}{2a}}}\log\left(\frac{t}{q}\right)^{\frac{1}{a}}\exp(-t)\prod_{j=1}^{p-1}B\left(\frac{p-j}{2},\frac{1}{2}\right)dt \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{a\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\left\{a\frac{\partial}{\partial N}\int_0^\infty t^{\frac{p+2N-2}{2a}-1}\exp(-t)-\Gamma\left(\frac{2N+p-2}{2a}\right)\log q\right\}\prod_{j=1}^{p-1}B\left(\frac{p-j}{2},\frac{1}{2}\right) \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{a\pi^{\frac{p}{2}}\Gamma\left(\frac{2N+p-2}{2a}\right)}\left\{a\frac{\partial}{\partial N}\Gamma\left(\frac{p+2N-2}{2a}\right)-\Gamma\left(\frac{2N+p-2}{2a}\right)\log q\right\}\prod_{j=1}^{p-1}B\left(\frac{p-j}{2},\frac{1}{2}\right) \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{a\pi^{\frac{p}{2}}}\left\{\psi\left(\frac{p+2N-2}{2a}\right)-\log q\right\}\prod_{j=1}^{p-1}B\left(\frac{p-j}{2},\frac{1}{2}\right),
\end{aligned}
$$

where $t=qr^{2a}$.

The I_2 can be calculated as

$$\begin{aligned}
I_2 &= \frac{\Gamma\left(\frac{p}{2}\right) q^{\frac{2N+p-2}{2a}}}{\pi^{\frac{p}{2}} \Gamma\left(\frac{2N+p-2}{2a}\right)} \int_0^\infty \frac{t^{\frac{2N+p-2}{2a}-1}}{q^{\frac{2N+p-2}{2a}}} \exp(-t) \\
&\quad \times \int_0^1 \cdots \int_0^1 \left\{ \prod_{j=1}^{p-1} \left[x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} \right] \right\} \log\left[\mathcal{B}_p(x_1, \ldots, x_{p-1})\right] dx_1 \cdots dx_{p-1} dt \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \int_0^1 \cdots \int_0^1 \left\{ \prod_{j=1}^{p-1} \left[x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} \right] \right\} \\
&\quad \times \left\{ \log \lambda_1 + \log\left[1 + \frac{(\lambda_2 - \lambda_1)x_1}{\lambda_1} + \cdots + \frac{(\lambda_p - \lambda_{p-1})x_1 x_2 \cdots x_{p-1}}{\lambda_1}\right] \right\} dx_1 \cdots dx_{p-1} \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \log \lambda_1 \left[\prod_{j=1}^{p-1} B\left(\frac{p-j}{2}, \frac{1}{2}\right) \right] - \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \int_0^1 \cdots \int_0^1 \left\{ \prod_{j=1}^{p-1} \left[x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} \right] \right\} \\
&\quad \times \sum_{k=1}^\infty \frac{(-1)^k}{k} \left[\frac{(\lambda_2 - \lambda_1)x_1}{\lambda_1} + \cdots + \frac{(\lambda_p - \lambda_{p-1})x_1 x_2 \cdots x_{p-1}}{\lambda_1} \right]^k dx_1 \cdots dx_{p-1} \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \log \lambda_1 \left[\prod_{j=1}^{p-1} B\left(\frac{p-j}{2}, \frac{1}{2}\right) \right] - \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \sum_{k=1}^\infty \frac{(-1)^k}{k} \sum_{i_1+\cdots+i_{p-1}=k} \binom{k}{i_1, \ldots, i_{p-1}} \\
&\quad \times \int_0^1 \cdots \int_0^1 \prod_{j=1}^{p-1} \left\{ \left[x_j^{\frac{p-j}{2}-1} (1-x_j)^{-\frac{1}{2}} \right] \left[\frac{(\lambda_{j+1} - \lambda_j)}{\lambda_1} \left(\prod_{\ell=1}^{j} x_\ell \right) \right]^{i_j} \right\} dx_1 \cdots dx_{p-1} \\
&= \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \log \lambda_1 \left[\prod_{j=1}^{p-1} B\left(\frac{p-j}{2}, \frac{1}{2}\right) \right] - \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \sum_{k=1}^\infty \frac{(-1)^k}{k} \sum_{i_1+\cdots+i_{p-1}=k} \binom{k}{i_1, \ldots, i_{p-1}} \\
&\quad \times \prod_{j=1}^{p-1} \left\{ \left[\frac{(\lambda_{j+1} - \lambda_j)}{\lambda_1} \right]^{i_j} B\left(\sum_{\ell=j}^{p-1} i_\ell + \frac{p-j}{2}, \frac{1}{2} \right) \right\}
\end{aligned}$$

provided that the infinite sum converges.

Hence, the required.

4.7. *Proof for Section 2.6*

The corresponding KLD can be expressed as

$$\begin{aligned}
KLD &= \log\left(\frac{a_1 \cdots a_p}{b_1 \cdots b_p}\right) + \sum_{\ell=1}^{p} (b_\ell - a_\ell) E(X_\ell) \\
&\quad + (p+1) E\left\{\log\left[1 + \exp(-b_1 X_1) + \cdots + \exp(-b_p X_p)\right]\right\} \\
&\quad - (p+1) E\left\{\log\left[1 + \exp(-a_1 X_1) + \cdots + \exp(-a_p X_p)\right]\right\} \\
&= \log\left(\frac{a_1 \cdots a_p}{b_1 \cdots b_p}\right) + (p+1) E\left\{\log\left[1 + \exp(-b_1 X_1) + \cdots + \exp(-b_p X_p)\right]\right\} \\
&\quad - (p+1) E\left\{\log\left[1 + \exp(-a_1 X_1) + \cdots + \exp(-a_p X_p)\right]\right\}
\end{aligned} \tag{11}$$

since the expectations are zero. Using the Taylor expansion for $\log(1+z)$, the first expectation in (11) can be expressed as

$$
\begin{aligned}
&-\sum_{k=1}^{\infty}\frac{(-1)^k}{k}E\Big\{\left[\exp(-b_1X_1)+\cdots+\exp(-b_pX_p)\right]^k\Big\}\\
&=-\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{i_1+\cdots+i_p=k}\binom{k}{i_1,\ldots,i_p}E\left[\exp(-i_1b_1X_1-\cdots-i_pb_pX_p)\right]\\
&=-\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{i_1+\cdots+i_p=k}\binom{k}{i_1,\ldots,i_p}\left[\prod_{j=1}^{p}\Gamma\left(\frac{a_j+i_jb_j}{a_j}\right)\right]\Gamma\left(1-\sum_{j=1}^{p}\frac{i_jb_j}{a_j}\right),
\end{aligned}
$$

where the last step follows by Lemma 1 provided that $\sum_{j=1}^{p}\frac{i_jb_j}{a_j}<1$ and the infinite series converges. The second expectation in (11) can be expressed as

$$
\begin{aligned}
&p!a_1\cdots a_p\int_{\mathbb{R}^p}\left\{\frac{\partial}{\partial\alpha}\left[1+\exp(-a_1x_1)+\cdots+\exp(-a_px_p)\right]^{\alpha}\Big|_{\alpha=0}\right\}\\
&\qquad\times\frac{\exp(-a_1x_1-\cdots-a_px_p)dx_1\cdots dx_p}{\left[1+\exp(-a_1x_1)+\cdots+\exp(-a_px_p)\right]^{p+1}}\\
&=p!a_1\cdots a_p\frac{\partial}{\partial\alpha}\left\{\int_{\mathbb{R}^p}\frac{\exp(-a_1x_1-\cdots-a_px_p)}{\left[1+\exp(-a_1x_1)+\cdots+\exp(-a_px_p)\right]^{p+1-\alpha}}dx_1\cdots dx_p\right\}\Bigg|_{\alpha=0}\\
&=p!\frac{\partial}{\partial\alpha}\left[\frac{\Gamma(1-\alpha)}{\Gamma(p+1-\alpha)}\right]\Bigg|_{\alpha=0}\\
&=\frac{\Gamma'(p+1)-\Gamma(p+1)\Gamma'(1)}{p!},
\end{aligned}
$$

where the penultimate step is followed by Lemma 1. Hence, the required.

4.8. Proof for Section 2.7

The corresponding KLD can be expressed as

$$
\begin{aligned}
KLD&=\log\left[\frac{(b)_pa_1\cdots a_p}{(d)_pb_1\cdots b_p}\right]+\sum_{\ell=1}^{p}(c_\ell-a_\ell)E(X_\ell)\\
&\quad+(d+p)E\left\{\log\left[1+\exp(-c_1X_1)+\cdots+\exp(-c_pX_p)\right]\right\}\\
&\quad-(b+p)E\left\{\log\left[1+\exp(-a_1X_1)+\cdots+\exp(-a_pX_p)\right]\right\}\\
&=\log\left[\frac{(b)_pa_1\cdots a_p}{(d)_pb_1\cdots b_p}\right]+(d+p)E\left\{\log\left[1+\exp(-c_1X_1)+\cdots+\exp(-c_pX_p)\right]\right\}\\
&\quad-(b+p)E\left\{\log\left[1+\exp(-a_1X_1)+\cdots+\exp(-a_pX_p)\right]\right\}
\end{aligned}
\tag{12}
$$

since the expectations are zero. Using the Taylor expansion for $\log(1+z)$, the first expectation in (12) can be expressed as

$$
\begin{aligned}
&-\sum_{k=1}^{\infty}\frac{(-1)^k}{k}E\Big\{\left[\exp(-c_1X_1)+\cdots+\exp(-c_pX_p)\right]^k\Big\}\\
&=-\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{i_1+\cdots+i_p=k}\binom{k}{i_1,\ldots,i_p}E\left[\exp(-i_1c_1X_1-\cdots-i_pc_pX_p)\right]\\
&=-\frac{1}{\Gamma(b)}\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{i_1+\cdots+i_p=k}\binom{k}{i_1,\ldots,i_p}\left[\prod_{j=1}^{p}\Gamma\left(\frac{a_j+i_jc_j}{a_j}\right)\right]\Gamma\left(b-\sum_{j=1}^{p}\frac{i_jc_j}{a_j}\right),
\end{aligned}
$$

where the last step follows by Lemma 1 provided that $\sum_{j=1}^{p} \frac{i_j c_j}{a_j} < b$ and the infinite series converges. The second expectation in (12) can be expressed as

$$
\begin{aligned}
&(b)_p a_1 \cdots a_p \int_{\mathbb{R}^p} \left\{ \frac{\partial}{\partial \alpha} \left[1 + \exp(-a_1 x_1) + \cdots + \exp(-a_p x_p) \right]^{\alpha} \Big|_{\alpha=0} \right\} \\
&\qquad \times \frac{\exp(-a_1 x_1 - \cdots - a_p x_p) dx_1 \cdots dx_p}{\left[1 + \exp(-a_1 x_1) + \cdots + \exp(-a_p x_p) \right]^{b+p}} \\
&= (b)_p a_1 \cdots a_p \frac{\partial}{\partial \alpha} \left\{ \int_{\mathbb{R}^p} \frac{\exp(-a_1 x_1 - \cdots - a_p x_p)}{\left[1 + \exp(-a_1 x_1) + \cdots + \exp(-a_p x_p) \right]^{b+p-\alpha}} dx_1 \cdots dx_p \right\} \Big|_{\alpha=0} \\
&= (b)_p \frac{\partial}{\partial \alpha} \left[\frac{\Gamma(b-\alpha)}{\Gamma(b+p-\alpha)} \right] \Big|_{\alpha=0} \\
&= \frac{\Gamma(b)\Gamma'(b+p) - \Gamma(b+p)\Gamma'(b)}{\Gamma(b)\Gamma(b+p)},
\end{aligned}
$$

where the penultimate step is followed by Lemma 1. Hence, the required.

4.9. Proof for Section 2.7

The corresponding KLD can be expressed as

$$
\begin{aligned}
KLD = \log \left[\frac{\sqrt{a_1 \cdots a_p} \beta_p(c, a_1, \ldots, a_p)}{\sqrt{b_1 \cdots b_p} \beta_p(d, b_1, \ldots, b_p)} \right] \\
+ \frac{1}{2} \sum_{i=1}^{p} (b_i - a_i) E\left(X_i^2 \right) \\
+ \frac{1}{2} (d b_1 \cdots b_p - c a_1 \cdots a_p) E\left(X_1^2 \cdots X_p^2 \right).
\end{aligned} \tag{13}
$$

Using results in [8], we can calculate (13) as the required.

4.10. Proof for Section 2.9

The corresponding KLD can be expressed as

$$
\begin{aligned}
KLD = \log \left[\frac{\Gamma\left(\frac{K}{2} + a + b - 1\right) \Gamma(c) \Gamma\left(\frac{K}{2} + d - 1\right)}{\Gamma(a) \Gamma\left(\frac{K}{2} + b - 1\right) \Gamma\left(\frac{K}{2} + c + d - 1\right)} \right] + (b-d) E\left[\log\left(\sum_{i=1}^{K} X_i^2 \right) \right] \\
+ (a-c) E\left[\log\left(1 - \sum_{i=1}^{K} X_i^2 \right) \right].
\end{aligned} \tag{14}
$$

It is easy to show that

$$
E\left[\log\left(\sum_{i=1}^{K} X_i^2 \right) \right] = \psi\left(\frac{K}{2} + b - 1 \right) - \psi\left(\frac{K}{2} + a + b - 1 \right)
$$

and

$$
E\left[\log\left(1 - \sum_{i=1}^{K} X_i^2 \right) \right] = \psi(a) - \psi\left(\frac{K}{2} + a + b - 1 \right),
$$

so (14) reduces to the required.

4.11. Proof for Section 2.10

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{C(d,e,f)}{C(a,b,c)}\right] + (a-d)E\left[\log\left(\prod_{i=1}^{p} X_i\right)\right] + (b-e)E\left[\log\left(\prod_{i=1}^{p}(1-X_i)\right)\right] + 2(c-f)E\left[\log\left(\prod_{1\le i<j\le p}(X_i - X_j)\right)\right]. \tag{15}$$

Easy calculations show that

$$E\left[\log\left(\prod_{i=1}^{p} X_i\right)\right] = \frac{\partial}{\partial\alpha}\left[\frac{C(a,b,c)}{C(a+\alpha,b,c)}\right]\bigg|_{\alpha=0},$$

$$E\left[\log\left(\prod_{i=1}^{p}(1-X_i)\right)\right] = \frac{\partial}{\partial\alpha}\left[\frac{C(a,b,c)}{C(a,b+\alpha,c)}\right]\bigg|_{\alpha=0}$$

and

$$E\left[\log\left(\prod_{1\le i<j\le p}(X_i - X_j)\right)\right] = \frac{\partial}{\partial\alpha}\left[\frac{C(a,b,c)}{C(a,b,c+\alpha)}\right]\bigg|_{\alpha=0},$$

so (7) reduces to the required.

4.12. Proof for Section 2.11

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{b_{p+1}\left(\sum_{i=1}^{p} a_i\right)\left(\prod_{i=1}^{p} a_i\right)}{a_{p+1}\left(\sum_{i=1}^{p} b_i\right)\left(\prod_{i=1}^{p} b_i\right)}\right] + \sum_{i=1}^{p}(b_i - a_i)E(X_i) + E\left[\log\left\{1-\exp\left[-a_{p+1}\min\left(X_1,\ldots,X_p\right)\right]\right\}\right] - E\left[\log\left\{1-\exp\left[-b_{p+1}\min\left(X_1,\ldots,X_p\right)\right]\right\}\right]. \tag{16}$$

Using the series expansion for $\log(1+z)$, we can express (16) as

$$KLD = \log\left[\frac{b_{p+1}\left(\sum_{i=1}^{p} a_i\right)\left(\prod_{i=1}^{p} a_i\right)}{a_{p+1}\left(\sum_{i=1}^{p} b_i\right)\left(\prod_{i=1}^{p} b_i\right)}\right] + \sum_{i=1}^{p}(b_i - a_i)E(X_i) - \sum_{k=1}^{\infty}\frac{E\left\{\exp\left[-ka_{p+1}\min\left(X_1,\ldots,X_p\right)\right]\right\}}{k} + \sum_{k=1}^{\infty}\frac{E\left\{\exp\left[-kb_{p+1}\min\left(X_1,\ldots,X_p\right)\right]\right\}}{k}.$$

Hence, the required.

4.13. Proof for Section 2.12

The corresponding KLD can be expressed as

$$\begin{aligned} KLD &= \log\left[\frac{\kappa_1^{\frac{p}{2}-1}}{\kappa_2^{\frac{p}{2}-1}}\frac{I_{\frac{p}{2}-1}(\kappa_2)}{I_{\frac{p}{2}-1}(\kappa_1)}\right] + \left(\kappa_1\boldsymbol{\mu}_1^T - \kappa_2\boldsymbol{\mu}_2^T\right)E(\mathbf{X}) \\ &= \log\left[\frac{\kappa_1^{\frac{p}{2}-1}}{\kappa_2^{\frac{p}{2}-1}}\frac{I_{\frac{p}{2}-1}(\kappa_2)}{I_{\frac{p}{2}-1}(\kappa_1)}\right] + \left(\kappa_1\boldsymbol{\mu}_1^T - \kappa_2\boldsymbol{\mu}_2^T\right)\boldsymbol{\mu}_1. \end{aligned}$$

Hence, the required.

4.14. Proof for Section 3.1

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{|\boldsymbol{\Omega}|^{c+d-a-b}B_p(c,d)}{B_p(a,b)}\right] + (a-c)E[\log|\boldsymbol{\Omega}-\mathbf{X}|] + (b-d)E[\log|\mathbf{X}|]. \tag{17}$$

The expectations in (17) can be calculated as

$$E[\log|\boldsymbol{\Omega}-\mathbf{X}|] = \frac{\partial}{\partial\alpha}\left\{\int \frac{|\boldsymbol{\Omega}-\mathbf{x}|^{\alpha+a-\frac{p+1}{2}}|\mathbf{x}|^{b-\frac{p+1}{2}}}{|\boldsymbol{\Omega}|^{a+b}B_p(a,b)}d\mathbf{x}\right\}\Bigg|_{\alpha=0} = \frac{\partial}{\partial\alpha}\left\{\frac{|\boldsymbol{\Omega}|^{\alpha}B_p(\alpha+a,b)}{B_p(a,b)}\right\}\Bigg|_{\alpha=0}$$

and

$$E[\log|\mathbf{X}|] = \frac{\partial}{\partial\alpha}\left\{\int \frac{|\boldsymbol{\Omega}-\mathbf{x}|^{a-\frac{p+1}{2}}|\mathbf{x}|^{\alpha+b-\frac{p+1}{2}}}{|\boldsymbol{\Omega}|^{a+b}B_p(a,b)}d\mathbf{x}\right\}\Bigg|_{\alpha=0} = \frac{\partial}{\partial\alpha}\left\{\frac{|\boldsymbol{\Omega}|^{\alpha}B_p(a,\alpha+b)}{B_p(a,b)}\right\}\Bigg|_{\alpha=0}.$$

Hence, the required.

4.15. Proof for Section 3.2

The corresponding KLD can be expressed as

$$\begin{aligned} KLD = \log&\left[\frac{B_p(b_1,\ldots,b_n;b_{n+1})}{B_p(a_1,\ldots,a_n;a_{n+1})}\right] + (a_{n+1}-b_{n+1})E\left[\log\left|\mathbf{I}_p - \sum_{i=1}^n \mathbf{X}_i\right|\right] \\ &+ \sum_{i=1}^n (a_i - b_i)E[\log|\mathbf{X}_i|]. \end{aligned} \tag{18}$$

It is easy to show that

$$E[\log|\mathbf{X}_i|] = \frac{\partial}{\partial\alpha}\frac{B_p(a_1,\ldots,a_i+\alpha,\ldots,a_n;a_{n+1})}{B_p(a_1,\ldots,a_i,\ldots,a_n;a_{n+1})}\Bigg|_{\alpha=0}$$

and

$$E\left[\log\left|\mathbf{I}_p - \sum_{i=1}^n \mathbf{X}_i\right|\right] = \frac{\partial}{\partial\alpha}\frac{B_p(a_1,\ldots,a_n;a_{n+1}+\alpha)}{B_p(a_1,\ldots,a_n;a_{n+1})}\Bigg|_{\alpha=0},$$

so (18) reduces to the required.

4.16. Proof for Section 3.3

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{d^{pc}\Gamma_p(c)}{b^{pa}\Gamma_p(a)}\frac{|\boldsymbol{\Sigma}_2|^c}{|\boldsymbol{\Sigma}_1|^a}\right] + (a-c)E[\log|\mathbf{X}|] + \mathrm{tr}\left[-\frac{1}{b}\boldsymbol{\Sigma}_1^{-1}E(\mathbf{X})\right] - \mathrm{tr}\left[-\frac{1}{d}\boldsymbol{\Sigma}_2^{-1}E(\mathbf{X})\right]. \quad (19)$$

The first expectation in (19) can be calculated as

$$\begin{aligned} E[\log|\mathbf{X}|] &= \frac{|\boldsymbol{\Sigma}_1|^{-a}}{b^{ap}\Gamma_p(a)}\frac{\partial}{\partial\alpha}\left\{\int |\mathbf{x}|^{\alpha+a-\frac{p+1}{2}}\exp\left[\mathrm{tr}\left(-\frac{1}{b}\boldsymbol{\Sigma}_1^{-1}\mathbf{x}\right)\right]\right\}\bigg|_{\alpha=0} \\ &= \frac{1}{\Gamma_p(a)}\frac{\partial}{\partial\alpha}\left[|\boldsymbol{\Sigma}_1|^{\alpha}b^{p\alpha}\Gamma_p(a+\alpha)\right]\bigg|_{\alpha=0}. \end{aligned}$$

Since $E(\mathbf{X}) = 2a\boldsymbol{\Sigma}_1$, the second and third terms in (19) are equal to

$$\mathrm{tr}\left[-\frac{1}{b}\boldsymbol{\Sigma}_1^{-1}E(\mathbf{X})\right] = \frac{2ap}{b}$$

and

$$\mathrm{tr}\left[-\frac{1}{d}\boldsymbol{\Sigma}_2^{-1}E(\mathbf{X})\right] = \mathrm{tr}\left[-\frac{2a}{d}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\right],$$

respectively. Hence, the required.

4.17. Proof for Section 3.4

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{B_p(d,e)\ {}_2F_1(d,f;d+e;-\mathbf{B})}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B})}\right] + (a-d)E[\log|\mathbf{X}|] + (b-e)E\left[\log\left|\mathbf{I}_p-\mathbf{X}\right|\right] + (f-c)E\left[\log\left|\mathbf{I}_p+\mathbf{B}\mathbf{X}\right|\right]. \quad (20)$$

The expectations in (20) can be easily calculated as

$$E[\log|\mathbf{X}|] = \frac{\partial}{\partial\alpha}\frac{B_p(a+\alpha,b)\ {}_2F_1(a+\alpha,c;a+b+\alpha;-\mathbf{B})}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B})}\bigg|_{\alpha=0},$$

$$E\left[\log\left|\mathbf{I}_p-\mathbf{X}\right|\right] = \frac{\partial}{\partial\alpha}\frac{B_p(a,b+\alpha)\ {}_2F_1(a,c;a+b+\alpha;-\mathbf{B})}{B_p(a,b)\ {}_2F_1(a,c;a+b;-\mathbf{B})}\bigg|_{\alpha=0}$$

and

$$E\left[\log\left|\mathbf{I}_p+\mathbf{B}\mathbf{X}\right|\right] = \frac{\partial}{\partial\alpha}\frac{{}_2F_1(a,c-\alpha;a+b;-\mathbf{B})}{{}_2F_1(a,c;a+b;-\mathbf{B})}\bigg|_{\alpha=0}.$$

Hence, the required.

4.18. Proof for Section 3.5

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{|\boldsymbol{\Omega}|^{a-c}B_p(c,d)}{B_p(a,b)}\right] + (c+d-a-b)E[\log|\boldsymbol{\Omega}+\mathbf{X}|] + (b-d)E[\log|\mathbf{X}|]. \quad (21)$$

The expectations in (21) can be calculated as

$$E[\log|\mathbf{\Omega}+\mathbf{X}|] = \frac{\partial}{\partial\alpha}\left\{\int \frac{|\mathbf{\Omega}+\mathbf{x}|^{\alpha-a-b}|\mathbf{x}|^{b-\frac{p+1}{2}}}{|\mathbf{\Omega}|^{-a}B_p(a,b)}d\mathbf{x}\right\}\Bigg|_{\alpha=0} = \frac{\partial}{\partial\alpha}\left\{\frac{|\mathbf{\Omega}|^{\alpha}B_p(a-\alpha,b)}{B_p(a,b)}\right\}\Bigg|_{\alpha=0}$$

and

$$E[\log|\mathbf{X}|] = \frac{\partial}{\partial\alpha}\left\{\int \frac{|\mathbf{\Omega}+\mathbf{x}|^{-a-b}|\mathbf{x}|^{\alpha+b-\frac{p+1}{2}}}{|\mathbf{\Omega}|^{-a}B_p(a,b)}d\mathbf{x}\right\}\Bigg|_{\alpha=0} = \frac{\partial}{\partial\alpha}\left\{\frac{|\mathbf{\Omega}|^{\alpha}B_p(a-\alpha,\alpha+b)}{B_p(a,b)}\right\}\Bigg|_{\alpha=0}.$$

Hence, the required.

4.19. Proof for Section 3.6

The corresponding KLD can be expressed as

$$\begin{aligned} KLD = {} & \log\left[\frac{d^{pc}\Gamma_p(c)}{b^{pa}\Gamma_p(a)}\frac{|\mathbf{\Sigma}_1|^a}{|\mathbf{\Sigma}_2|^c}\right] + (c-a)E[\log|\mathbf{X}|] + \frac{1}{d}\mathrm{tr}\left[\mathbf{\Sigma}_2 E\left(\mathbf{X}^{-1}\right)\right] \\ & - \frac{1}{b}\mathrm{tr}\left[\mathbf{\Sigma}_1 E\left(\mathbf{X}^{-1}\right)\right]. \end{aligned} \tag{22}$$

The first expectation in (22) can be calculated as

$$\begin{aligned} E[\log|\mathbf{X}|] &= \frac{|\mathbf{\Sigma}_1|^a}{b^{ap}\Gamma_p(a)}\frac{\partial}{\partial\alpha}\left\{\int |\mathbf{x}|^{\alpha-a-\frac{p+1}{2}}\exp\left[\mathrm{tr}\left(-\frac{1}{b}\mathbf{\Sigma}_1\mathbf{x}\right)\right]\right\}\Bigg|_{\alpha=0} \\ &= \frac{1}{\Gamma_p(a)}\frac{\partial}{\partial\alpha}\left[\frac{|\mathbf{\Sigma}_1|^{\alpha}\Gamma_p(a-\alpha)}{b^{p\alpha}}\right]\Bigg|_{\alpha=0}. \end{aligned}$$

Since $E\left(\mathbf{X}^{-1}\right) = 2a\mathbf{\Sigma}_1^{-1}$, the second and third terms in (22) are equal to

$$\mathrm{tr}\left[\mathbf{\Sigma}_2 E\left(\mathbf{X}^{-1}\right)\right] = 2a\ \mathrm{tr}\left[\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1}\right]$$

and

$$\mathrm{tr}\left[\mathbf{\Sigma}_1 E\left(\mathbf{X}^{-1}\right)\right] = 2ap,$$

respectively. Hence, the required.

4.20. Proof for Section 3.7

The corresponding KLD can be expressed as

$$\begin{aligned} KLD = {} & \log\left[\frac{B_p(c,d)\ {}_1F_1(c;c+d;-\mathbf{B}_2)}{B_p(a,b)\ {}_1F_1(a,a+b;-\mathbf{B}_1)}\right] + (a-c)E[\log|\mathbf{X}|] + (b-d)E\left[\log|\mathbf{I}_p-\mathbf{X}|\right] \\ & + \mathrm{tr}\left[(\mathbf{B}_2-\mathbf{B}_1)E(\mathbf{X})\right]. \end{aligned} \tag{23}$$

The expectations in (23) can be easily calculated as

$$E[\log|\mathbf{X}|] = \frac{\partial}{\partial\alpha}\frac{B_p(a+\alpha,b)\ {}_1F_1(a+\alpha;a+b+\alpha;-\mathbf{B}_1)}{B_p(a,b)\ {}_1F_1(a;a+b;-\mathbf{B}_1)}\Bigg|_{\alpha=0},$$

$$E\left[\log|\mathbf{I}_p-\mathbf{X}|\right] = \frac{\partial}{\partial\alpha}\frac{B_p(a,b+\alpha)\ {}_1F_1(a;a+b+\alpha;-\mathbf{B}_1)}{B_p(a,b)\ {}_1F_1(a;a+b;-\mathbf{B}_1)}\Bigg|_{\alpha=0}$$

and

$$E(\mathbf{X}) = \frac{\partial}{\partial \mathbf{z}} \frac{{}_1F_1(a; a+b; \mathbf{z}-\mathbf{B}_1)}{{}_1F_1(a; a+b; -\mathbf{B}_1)}\bigg|_{\mathbf{z}=\mathbf{0}}.$$

Hence, the required.

4.21. Proof for Section 3.8

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{\Gamma_p(c)\,\Psi_1\left(c; c-d+\frac{p+1}{2}; \mathbf{B}_2\right)}{\Gamma_p(a)\,\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\right] + (a-c)E[\log|\mathbf{X}|] + (d-b)E\left[\log\left|\mathbf{I}_p+\mathbf{X}\right|\right] + \operatorname{tr}\left[(\mathbf{B}_2-\mathbf{B}_1)E(\mathbf{X})\right]. \tag{24}$$

The expectations in (24) can be easily calculated as

$$E[\log|\mathbf{X}|] = \frac{\partial}{\partial \alpha} \frac{\Gamma_p(a+\alpha)\,\Psi_1\left(a+\alpha; a-b+\alpha+\frac{p+1}{2}; \mathbf{B}_1\right)}{\Gamma_p(a)\,\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\Bigg|_{\alpha=0},$$

$$E\left[\log\left|\mathbf{I}_p+\mathbf{X}\right|\right] = \frac{\partial}{\partial \alpha} \frac{\Psi_1\left(a; a-b+\alpha+\frac{p+1}{2}; \mathbf{B}_1\right)}{\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\Bigg|_{\alpha=0}$$

and

$$E(\mathbf{X}) = \frac{\partial}{\partial \mathbf{z}} \frac{\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1-\mathbf{z}\right)}{\Psi_1\left(a; a-b+\frac{p+1}{2}; \mathbf{B}_1\right)}\Bigg|_{\mathbf{z}=\mathbf{0}}.$$

Hence, the required.

4.22. Proof for Section 3.9

The corresponding KLD can be expressed as

$$KLD = \log\left[\frac{|\mathbf{V}_2|^{\frac{n}{2}}|\mathbf{U}_2|^{\frac{p}{2}}}{|\mathbf{V}_1|^{\frac{n}{2}}|\mathbf{U}_1|^{\frac{p}{2}}}\right] + \frac{1}{2}E\left\{\operatorname{tr}\left[\mathbf{V}_2^{-1}(\mathbf{X}-\mathbf{M}_2)^T\mathbf{U}_2^{-1}(\mathbf{X}-\mathbf{M}_2)\right]\right\} - \frac{1}{2}E\left\{\operatorname{tr}\left[\mathbf{V}_1^{-1}(\mathbf{X}-\mathbf{M}_1)^T\mathbf{U}_1^{-1}(\mathbf{X}-\mathbf{M}_1)\right]\right\}. \tag{25}$$

The second expectation in (25) can be expressed as

$$\operatorname{tr}\left\{\mathbf{V}_1^{-1}E\left[(\mathbf{X}-\mathbf{M}_1)^T(\mathbf{X}-\mathbf{M}_1)\right]\mathbf{U}_1^{-1}\right\} = \operatorname{tr}(\mathbf{U}_1)\operatorname{tr}\left(\mathbf{U}_1^{-1}\right).$$

The first expectation in (25) can be expressed as

$$\begin{aligned}
&\operatorname{tr}\left\{\mathbf{V}_2^{-1}E\left[(\mathbf{X}-\mathbf{M}_2)^T(\mathbf{X}-\mathbf{M}_2)\right]\mathbf{U}_2^{-1}\right\} \\
&= \operatorname{tr}\left[\mathbf{V}_2^{-1}E\left(\mathbf{X}^T\mathbf{X}-\mathbf{X}^T\mathbf{M}_2-\mathbf{M}_2^T\mathbf{X}+\mathbf{M}_2^T\mathbf{M}_2\right)\mathbf{U}_2^{-1}\right] \\
&= \operatorname{tr}\left\{\mathbf{V}_2^{-1}\left[E\left(\mathbf{X}^T\mathbf{X}\right)-E\left(\mathbf{X}^T\mathbf{M}_2\right)-E\left(\mathbf{M}_2^T\mathbf{X}\right)+\mathbf{M}_2^T\mathbf{M}_2\right]\mathbf{U}_2^{-1}\right\} \\
&= \operatorname{tr}(\mathbf{U}_1)\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{V}_1\mathbf{U}_2^{-1}\right)+\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_1^T\mathbf{M}_1\mathbf{U}_2^{-1}\right)-\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_1^T\mathbf{M}_2\mathbf{U}_2^{-1}\right) \\
&\quad -\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_2^T\mathbf{M}_1\mathbf{U}_2^{-1}\right)+\operatorname{tr}\left(\mathbf{V}_2^{-1}\mathbf{M}_2^T\mathbf{M}_2\mathbf{U}_2^{-1}\right).
\end{aligned}$$

Hence, the required.

4.23. Proof for Section 3.10

The corresponding KLD can be expressed as

$$\begin{aligned}KLD = \log\left[\frac{C(a)}{C(b)}\right] &+ (a-b)E\left[\log|\mathbf{X}|I\{\mathbf{0}_p < \mathbf{X} \leq \mathbf{B}\}\right] \\ &+ (a-b)E\left[\log\left|\mathbf{I}_p - \mathbf{X}\right|I\{\mathbf{B} < \mathbf{X} \leq \mathbf{I}_p\}\right] \\ &+ (b-a)|\mathbf{B}|^{\frac{p+1}{2}}\log|\mathbf{B}|B_p\left(a, \frac{p+1}{2}\right) \\ &+ (b-a)\left|\mathbf{I}_p - \mathbf{B}\right|^{\frac{p+1}{2}}\log\left|\mathbf{I}_p - \mathbf{B}\right|B_p\left(\frac{p+1}{2}, a\right). \end{aligned} \quad (26)$$

The expectations in (26) can be easily calculated as

$$E\left[\log|\mathbf{X}|I\{\mathbf{0}_p < \mathbf{X} \leq \mathbf{B}\}\right] = \frac{\partial}{\partial\alpha}\left\{|\mathbf{B}|^{\alpha+\frac{p+1}{2}}B_p\left(a+\alpha, \frac{p+1}{2}\right)\right\}\Bigg|_{\alpha=0}$$

and

$$E\left[\log\left|\mathbf{I}_p - \mathbf{X}\right|I\{\mathbf{B} < \mathbf{X} \leq \mathbf{I}_p\}\right] = \frac{\partial}{\partial\alpha}\left\{\left|\mathbf{I}_p - \mathbf{B}\right|^{\alpha+\frac{p+1}{2}}B_p\left(\frac{p+1}{2}, a+\alpha\right)\right\}\Bigg|_{\alpha=0}.$$

Hence, the required.

Author Contributions: Conceptualization, V.N. and S.N.; methodology, V.N. and S.N.; investigation, V.N. and S.N. All authors have read and agreed to the published version of the manuscript.

Funding: Research of the first author was partially supported by grants from the IMU-CDC, Simons Foundation, and the Heilbronn Institute for Mathematical Research.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to thank the Editor and the two referees for careful reading and comments which greatly improved the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Special Functions

The following special functions are used in the paper: the gamma function defined by

$$\Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)dt$$

for $a > 0$; the digamma function is defined by

$$\psi(a) = \frac{d\log\Gamma(a)}{da}$$

for $a > 0$; the beta function is defined by

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$$

for $a > 0$ and $b > 0$; the type I Dirichlet density is defined by

$$B(a_1, \ldots, a_n; a_{n+1}) = \int_{0\leq t_1+\cdots+t_n\leq 1} t_1^{a_1-1}\cdots t_n^{a_n-1}\left(1 - \sum_{i=1}^n t_i\right)^{a_{n+1}-1} dt_1\cdots dt_n$$

for $a_j > 0$, $j = 1, 2, \ldots, k+1$; the modified Bessel function of the first kind of order ν defined by

$$I_\nu(x) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+\nu+1)k!}\left(\frac{x}{2}\right)^{2k+\nu}$$

for $k+\nu+1 \neq 0, -1, -2, \ldots$; the matrix-variate gamma function is defined by

$$\Gamma_p(\alpha) = \int |\mathbf{x}|^{\alpha - \frac{p+1}{2}} \exp[-\mathrm{tr}\,(\mathbf{x})] d\mathbf{x}$$

for $\mathbf{x}$ a $p \times p$ positive definite matrix and $a > \frac{p-1}{2}$; the matrix-variate beta function is defined by

$$B_p(\alpha, \beta) = \int \left|\mathbf{I}_p - \mathbf{x}\right|^{\alpha - \frac{p+1}{2}} |\mathbf{x}|^{\beta - \frac{p+1}{2}} d\mathbf{x}$$

for $\mathbf{x}$ a $p \times p$ positive definite matrix, $\mathbf{I}_p - \mathbf{x}$ a $p \times p$ positive definite matrix, $\alpha > \frac{p-1}{2}$ and $\beta > \frac{p-1}{2}$; the matrix-variate type I Dirichlet density is defined by

$$B_p(a_1, \ldots, a_n; a_{n+1}) = \int |\mathbf{x}_1|^{a_1 - p} \cdots |\mathbf{x}_n|^{a_n - p} \left|\mathbf{I}_p - \sum_{i=1}^{n} \mathbf{x}_i\right|^{a_{n+1} - p} d\mathbf{x}_1 \cdots d\mathbf{x}_n$$

for $\mathbf{x}_i$, $i = 1, 2, \ldots, n$ and $\mathbf{I}_p - \sum_{i=1}^{n} \mathbf{x}_i$ being $p \times p$ positive definite matrices, and $a_j > \frac{p-1}{2}$, $j = 1, 2, \ldots, n+1$; the matrix-variate confluent hypergeometric function is defined by

$${}_1F_1(a; b; \mathbf{X}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a)_\kappa}{(b)_\kappa} \frac{C_\kappa(\mathbf{X})}{k!}$$

for $\mathbf{X}$ a $p \times p$ positive definite matrix and provided that $\Gamma_p(\alpha)$ and $\Gamma_p(\beta)$ exist; the matrix-variate Gauss hypergeometric function is defined by

$${}_2F_1(a, b; c; \mathbf{X}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a)_\kappa (b)_\kappa}{(c)_\kappa} \frac{C_\kappa(\mathbf{X})}{k!}$$

for $\mathbf{X}$ a $p \times p$ positive definite matrix and provided that $\Gamma_p(a)$, $\Gamma_p(b)$ and $\Gamma_p(c)$ exist, where $C_\kappa(\mathbf{X})$ denotes the zonal polynomial of the $p \times p$ symmetric matrix $\mathbf{X}$ corresponding to the ordered partition $\kappa = (k_1, \ldots, k_p)$ with $k_1 \geq \cdots \geq k_p \geq 0$ and $k_1 + \cdots + k_p = k$, $\sum_\kappa$ denotes summation over all such partitions κ, and

$$(a)_\kappa = \prod_{i=1}^{p} \left(a - \frac{i-1}{2}\right)_{k_i},$$

where $(a)_0 = 1$ and $(a)_k = a(a+1) \cdot (a+k-1)$.

References

1. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
2. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Cham, Switzerland, 2006.
3. Bouhlel, N.; Dziri, A. Kullback–Leibler divergence between multivariate generalized gaussian distributions. *IEEE Signal Process. Lett.* **2019**, *26*, 1021–1025. [CrossRef]
4. 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. [CrossRef] [PubMed]
5. Bouhlel, N.; Rousseau, D. Exact Rényi and Kullback–Leibler divergences between multivariate *t*-distributions. *IEEE Signal Process. Lett.* **2023**, *30*, 1672–1676. [CrossRef]
6. Malik, H.J.; Abraham, B. Multivariate logistic distributions. *Ann. Stat.* **1973**, *1*, 588–590. [CrossRef]

7. Satterthwaite, S.P.; Hutchinson, T.P. A generalization of Gumbel's bivariate logistic distribution. *Metrika* **1978**, *25*, 163–170. [CrossRef]
8. Sarabia, J.-M. The centered normal conditional distributions. *Commun. Stat.-Theory Methods* **1995**, *24*, 2889–2900. [CrossRef]
9. Penny, W.D. *Kullback-Liebler Divergences of Normal, Gamma, Dirichlet and Wishart Densities*; Wellcome Department of Cognitive Neurology: London, UK, 2001.
10. Kotz, S.; Balakrishnan, N.; Johnson, N.L. *Continuous Multivariate Distributions*; John Wiley and Sons: New York, NY, USA, 2000.
11. Nagar, D.K.; Bedoya-Valencia, D.; Nadarajah, S. Multivariate generalization of the Gauss hypergeometric distribution. *Hacet. J. Math. Stat.* **2015**, *44*, 933–948. [CrossRef]
12. Kotz, S. Multivariate distributions at a cross-road. In *Statistical Distributions in Scientific Work*; Patil, G.P., Kotz, S., Ord, J.K., Eds.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1975; Volume 1, pp. 247–270.
13. Pham-Gia, T. The multivariate Selberg beta distribution and applications. *Statistics* **2009**, *43*, 65–79. [CrossRef]
14. Al-Mutairi, D.K.; Ghitany, M.E.; Kundu, D. A new bivariate distribution with weighted exponential marginals and its multivariate generalization. *Stat. Pap.* **2011**, *52*, 921–936. [CrossRef]
15. Dawid, A.P. Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika* **1981**, *68*, 265–274. [CrossRef]
16. Gupta, A.K.; Nagar, D.K. Matrix-variate Gauss hypergeometric distribution. *J. Aust. Math. Soc.* **2012**, *92*, 335–355. [CrossRef]
17. Nagar, D.K.; Gupta, A.K. Matrix-variate Kummer-beta distribution. *J. Aust. Math. Soc.* **2002**, *73*, 11–25. [CrossRef]
18. Nagar, D.K.; Cardeno, L. Matrix variate Kummer-gamma distribution. *Random Oper. Stoch. Equ.* **2001**, *9*, 207–218.
19. Zinodiny, S.; Nadarajah, S. Matrix variate two-sided power distribution. *Methodol. Comput. Appl. Probab.* **2022**, *24*, 179–194. [CrossRef]