PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 46, Fasc. 1 (2026), pp. 105–125
Published online 12.10.2026
doi:10.37190/0208-4147.00308

UPPER AND LOWER BOUNDS ON TVD AND KLD BETWEEN CENTERED ELLIPTICAL DISTRIBUTIONS IN HIGH-DIMENSIONAL SETTINGS*

BY

PAVEL IEVLEV (LAUSANNE) AND TIMOFEI SHASHKOV (LAUSANNE)

Abstract. We derive upper and lower bounds for the total variation distance (TVD) and the Kullback–Leibler divergence (KLD), also known as the relative entropy, between two probability measures μ and ν defined by

$$D_{\mathrm{TV}}(\mu,\nu) = \sup_{B\in\mathcal{B}(\mathbb{R}^n)} |\mu(B)-\nu(B)|$$

and

$$D_{\mathrm{KL}}(\mu\,\|\,\nu) = \int\limits_{\mathbb{R}^n} \ln\left(\frac{d\mu(x)}{d\nu(x)}\right)\mu(dx)$$

in high-dimensional settings. We establish elementary bounds for centered elliptical distributions and apply them to multivariate Student t-distributions versus multivariate normal distributions. For the latter case, we provide an exact formula via integration over the unit sphere, computable bounds in terms of special functions, and concentration-based bounds in which the worst-case parameters are replaced by effective averaged quantities, up to exponentially small exceptional-set corrections. We also apply our approach to multivariate Gamma distributions with independent components. Our work is motivated by the recent contribution of Barabesi and Pratelli (2024).

2020 Mathematics Subject Classification: Primary 60E15; Secondary 62H05.

Key words and phrases: total variation distance, Kullback–Leibler divergence, multivariate Student distribution, high-dimensional statistics, Gamma distribution.

1. INTRODUCTION

Measuring the dissimilarity between two probability measures using total variation distance (TVD) or Kullback–Leibler divergence (KLD) is prominent across statistics and machine learning. In a typical application, the two measures are either (a) empirical and fitted, (b) corresponding to two different models of the

* Research supported by SNSF Grant 200021–196888.

© Probability and Mathematical Statistics, 2026

same dataset or (c) corresponding to one model fitted to two different datasets. In case (a), these measures are used to estimate how well the model fits the data. For example, in hypothesis testing TVD provides so-called minimax lower bounds on both Type I errors (false positives) and Type II errors (false negatives) over all possible decision rules [24, Theorem 2.2]. In case (b), the two measures assess whether a model may be switched to a more preferable one within prescribed error bounds. A common example is to switch a true but intractable distribution μ to a simple approximation ν, such as the Laplace approximation [5, 13, 22, 23] or variational inference [3]. In case (c), measures D_{TV} and D_{KL} are used to guarantee some robustness under changes of the underlying data, such as restriction to a subset, and/or mitigate overfitting. Since the advent of generative neural networks, case (c) has gained a lot of attention. Two notable examples for (c) are (i) assessing the vulnerability to data poisoning by estimating how much the fitted distribution is affected by injecting new points into the dataset [15] and (ii) providing quantitative privacy guarantee when the dataset used for training contains sensitive information, such as a language model trained on private emails. In both examples, the framework is quite similar: given a randomized algorithm assigning to a dataset $\mathcal{D}$ some distribution $\mu_{\mathcal{D}}$, check how $\mu_{\mathcal{D}}$ differs from $\mu_{\mathcal{D}'}$ if $\mathcal{D}$ and $\mathcal{D}'$ are close. In the differential privacy setting, the aim is to ensure that an observer cannot infer whether a particular individual's data was part of the training set by sampling from $\mu_{\mathcal{D}}$ by adding as little noise to the model as possible (utility-privacy trade-off) [12, 19]. In this context, TVD measures the maximum distance between the output distributions when a single data point is removed.

The choice between D_{KL} and D_{TV} in specific applications is mainly guided by (a) the necessity for a more sensitive metric (D_{TV} is weaker than D_{KL} due to Pinsker's inequality $D_{\mathrm{TV}}(\mu,\nu)^2 \leqslant D_{\mathrm{KL}}(\mu \,\|\, \nu)/2$) [11, 8, 25], (b) the fact that D_{KL} is frequently easier to compute (for many common distribution families, closed form expressions are available [18]), and (c) the fact that D_{TV} is usually easier to interpret. The sensitivity (a) is particularly important in the context of data poisoning and differential privacy [11, 8] whereas for more generic learning algorithms performance (b) is of more importance. For a general overview of probability metrics and how to choose between them, see [9].

The aforementioned applications underscore the necessity of total variation and KL divergence bounds *in high-dimensional settings* where explicit formulas are seldom available and Monte Carlo simulations may be prohibitively costly. This area of research has attracted a lot of attention recently [6, 2].

In this paper, we derive some elementary bounds on the TVD and KLD between centered elliptical distributions (Proposition 2.1) and then show how they may be applied when one distribution is the multivariate Student t-distribution and the other is multivariate normal (Theorems 3.2 and 3.4). We also provide an exact formula for the TVD in this case (Theorem 3.1), which requires integration over the unit sphere. The bounds and the exact formula are given in terms of standard special functions, such as the incomplete regularized Gamma and Beta functions

and the Lambert W-function. Since the bounds work for sufficiently large n, we show how to choose the threshold (see Lemma 3.3 and condition (3.7)). We also show how concentration of measure on the sphere yields bounds that replace the worst-case parameters d_- and d_+ by effective averaged quantities, up to exponentially small exceptional-set corrections (Theorem 3.3). Additionally, we show how a similar approach supplemented by an application of CLT with Berry–Esseen bounds with respect to n yields bounds on the TVD between two multivariate Gamma laws with independent components (Theorem 4.1). The latter idea has recently been used in [2].

In principle, the same approach is applicable to any two elliptical distributions for which the sets $A_\pm$ from Proposition 2.1 may be described in a convenient form. Moreover, the same approach may be used to construct bounds on other divergences.

Organization of the paper. Section 2 establishes elementary bounds on the TVD and KLD between centered elliptical distributions. Section 3 applies these to multivariate Student versus multivariate normal distributions, deriving an exact formula, computable bounds in terms of special functions, and concentration-based bounds based on effective averaged parameters. Section 4 treats multivariate Gamma distributions with independent components. Complete proofs are in the Appendix.

Notation. Throughout the paper, we use the following special functions:

- W_0 and W_{-1}: branches of the Lambert-W function [4];
- $\gamma(x;a) = \int_0^x t^{a-1}e^{-t}\,dt$ and $\Gamma(x;a) = \int_x^\infty t^{a-1}e^{-t}\,dt$: lower and upper incomplete Gamma function (see, e.g., [1, Chapter 6.5]);
- $P(z;a) = \gamma(x;a)/\Gamma(a)$: lower regularized incomplete Gamma function;
- $B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\,dt$: incomplete Beta function;
- $I(x;a,b) = B(x;a,b)/B(a,b)$: regularized incomplete Beta function;
- ${}_2F_1$ and ${}_2F_2$: hypergeometric functions;
- $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ and $\operatorname{erfi}(x) = -i\operatorname{erf}(x)$: error function and the imaginary error function;
- $\operatorname{Ei}(x) = \int_{-\infty}^x \frac{e^t}{t}\,dt$: exponential integral;
- $\psi_0(x) = \frac{d}{dx}\ln\Gamma(x)$ and $\psi_1(x) = \frac{d^2}{dx^2}\ln\Gamma(x)$: digamma and trigamma functions (polygamma functions of order 0 and 1).

References concerning these special functions may be found in [16].

2. ELEMENTARY BOUNDS ON TVD AND KLD BETWEEN TWO ELLIPTICAL DISTRIBUTIONS

Let $\mathcal{E}(g,\Sigma)$ be a centered elliptical distribution admitting a pdf given by

$$p(x) = \frac{1}{\sqrt{\det\Sigma}}g(x^\top\Sigma^{-1}x).$$

Here g is a so-called density generator. Unless specified otherwise, we assume in this section that g is a positive, measurable function such that p is a true probability density function. For a comprehensive treatment of elliptical distributions, see [7, 10].

First, we want to show that calculation of the distances between two elliptical distributions $\mathcal{E}(g_1, \Sigma_1)$ and $\mathcal{E}(g_2, \Sigma_2)$ may be reduced to the case when $\Sigma_1 = I$ and Σ_2 is diagonal. Here I is the identity matrix and g_1, g_2 are the corresponding density generators with the same assumptions as for the function g. Instead of proving it for TV and DKL separately, we prove it for all f-divergences (see [21]) in Lemma 2.1. Note that a version of this lemma in a general context is available in [17].

LEMMA 2.1. *Let D be the eigenvalue matrix of $\Sigma_1^{-1}\Sigma_2$. Then*

$$D_f(\mathcal{E}(g_1, \Sigma_1) \,\|\, \mathcal{E}(g_2, \Sigma_2)) = D_f(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D)),$$

where D_f is the f-divergence associated to a function $f\colon [0, \infty) \to (-\infty, \infty]$ such that the integral

$$D_f(\mu \,\|\, \nu) = \int\limits_{\mathbb{R}^n} f\left(\frac{d\mu(x)}{d\nu(x)}\right) d\nu(x)$$

is well-defined.

Applying this lemma with $f(t) = t \ln t$ and $f(t) = |t - 1|/2$ yields the same claim for D_{KL} and D_{TV}.

Next, we shall use the following formula for the total variation distance in terms of the pdfs: if $\boldsymbol{X} \sim \mathcal{E}(g_1, I)$ has pdf p_1 and $\boldsymbol{Y} \sim \mathcal{E}(g_2, D)$ has pdf p_2, then

$$D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) = \mathbb{P}\{p_1(\boldsymbol{X}) \geqslant p_2(\boldsymbol{X})\} - \mathbb{P}\{p_1(\boldsymbol{Y}) \geqslant p_2(\boldsymbol{Y})\}. \tag{2.1}$$

Hereafter we assume that $d_i = D_{ii}$, $i = 1, \ldots, n$, and there exist positive constants d_-, d_+ such that $d_- \leqslant d_i \leqslant d_+$ for all $i = 1, \ldots, n$.

The following two lemmas contain elementary bounds on TVD and DKL between elliptical distributions.

PROPOSITION 2.1. *Let $\boldsymbol{X} \sim \mathcal{E}(g_1, I)$ and $\boldsymbol{Y} \sim \mathcal{E}(g_2, D)$.*

(a) *If $D = I$, then*

$$\begin{aligned} D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, I)) &= \frac{2\pi^{n/2}}{\Gamma(n/2)} \int\limits_0^\infty \mathbf{1}\{r^2 \in A\}(g_1(r^2) - g_2(r^2)) r^{n-1}\, dr \\ &= \frac{2\pi^{n/2}}{\Gamma(n/2)} \int\limits_0^\infty \mathbf{1}\{r^2 \in A^c\}(g_2(r^2) - g_1(r^2)) r^{n-1}\, dr \end{aligned}$$

with $A = \{t \in [0,\infty)\colon g_1(t) \geqslant g_2(t)\}$ *and*

$$D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, I)) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \ln\left(\frac{g_1(r^2)}{g_2(r^2)}\right) g_1(r^2)\, r^{n-1}\, dr.$$

(b) *If* g_1 *and* g_2 *are decreasing on* $[0,\infty)$*, then denoting*

$$V_\pm = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty [g_1(r^2)\, \mathbf{1}\{r^2 \in A_\pm\} - g_2(r^2)\, \mathbf{1}\{d_\pm r^2 \in A_\mp\}] r^{n-1}\, dr,$$

we have $V_- \leqslant D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) \leqslant V_+$ *with*

$$A_\pm = \left\{ t \in [0,\infty) : \prod_{i=1}^n d_i^{1/2} g_1(t) \geqslant g_2\left(\frac{t}{d_\mp}\right)\right\}, \tag{2.2}$$

and if g_2 *is decreasing on* $[0,\infty)$*, then denoting*

$$K_\pm = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty g_1(r^2) \ln\left(\frac{g_1(r^2)}{g_2(r^2/d_\mp)}\right) r^{n-1}\, dr + \frac{1}{2}\sum_{i=1}^n \ln d_i,$$

we have $K_- \leqslant D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D)) \leqslant K_+$.

3. MULTIVARIATE STUDENT T-DISTRIBUTION AGAINST MULTIVARIATE NORMAL DISTRIBUTION

The pdfs of the centered multivariate t-distribution with $\Sigma = I$ and of the multivariate normal distribution with $\Sigma = D = \mathrm{diag}(d_i)$ are given by

$$\begin{aligned} p_1(x) &= \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\, \nu^{n/2}\pi^{n/2}} \left[1 + \frac{1}{\nu}\sum_{i=1}^n x_i^2\right]^{-(\nu+n)/2} \\ p_2(x) &= \frac{1}{(2\pi)^{n/2}\prod_{i=1}^n d_i^{1/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{x_i^2}{d_i}\right). \end{aligned} \tag{3.1}$$

In this section, we apply Proposition 2.1 to obtain lower and upper bounds on the total variation distance and KL divergence between these two distributions.

3.1. Total variation distance. To apply Proposition 2.1, we need first to find a convenient description of the sets

$$A_\pm = \left\{ t \in [0,\infty) : \begin{array}{r} \left(\prod_{i=1}^n d_i\right)^{1/2} \dfrac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\, \nu^{n/2}\pi^{n/2}} \left[1 + \dfrac{t}{\nu}\right]^{-(\nu+n)/2} \\ \geqslant \dfrac{1}{(2\pi)^{n/2}} e^{-\frac{t}{2d_\mp}} \end{array} \right\}.$$

Rewriting $A_\pm$ as

$$A_\pm = \left\{ t \in [0,\infty) : \begin{array}{r} \left(1+\dfrac{t}{\nu}\right)^{(\nu+n)/2} \exp\left(-\dfrac{t}{2d_\mp}\right) \\ \leqslant \dfrac{\Gamma((\nu+n)/2)\cdot(2\pi)^{n/2}}{\Gamma(\nu/2)\nu^{n/2}\pi^{n/2}} \prod\limits_{i=1}^{n} d_i^{1/2} \end{array} \right\}$$

and denoting

(3.2)
$$c(n) = \frac{\Gamma((\nu+n)/2)\cdot(2\pi)^{n/2}}{\Gamma(\nu/2)\nu^{n/2}\pi^{n/2}} \prod_{i=1}^{n} d_i^{1/2} = \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)} \left(\frac{2}{\nu}\right)^{n/2} \prod_{i=1}^{n} d_i^{1/2},$$

we find that both A_+ and A_- are sublevel sets of the function $\varphi_{\alpha,\gamma}(x) = (1+x)^\alpha e^{-\gamma x}$:

(3.3)
$$\begin{aligned} &A_\pm = \left\{ t \in [0,\infty) : \varphi_{\alpha,\gamma_\pm}\left(\frac{t}{\nu}\right) \leqslant c(n) \right\}, \quad \text{where} \\ &\alpha(n) = \frac{\nu+n}{2}, \quad \gamma_\pm = \frac{\nu}{2d_\mp}. \end{aligned}$$

In the following lemma we simplify the description of $A_\pm$.

LEMMA 3.1. *Let $\varphi_{\alpha,\gamma}(x) = (1+x)^\alpha e^{-\gamma x}$ and α, γ, $c > 0$, and denote $D_{\alpha,\gamma}(c) = \{x \geqslant 0 \colon \varphi_{\alpha,\gamma}(x) \leqslant c\}$. If $\alpha \leqslant \gamma$, then $D_{\alpha,\gamma}(c) = [0,\infty)$ for $c \geqslant 1$, while $D_{\alpha,\gamma}(c) = [b_{\alpha,\gamma}(c),\infty)$ for $0 < c < 1$, where $b_{\alpha,\gamma}(c) > 0$.*

If $\alpha > \gamma$, let $c^* = e^\gamma(\alpha/\gamma e)^\alpha$. Then

$$D_{\alpha,\gamma}(c) = \begin{cases} [b_{\alpha,\gamma}(c),\infty), & 0 < c < 1, \\ \{0\} \cup [b_{\alpha,\gamma}(1),\infty), & c = 1, \\ [0,a_{\alpha,\gamma}(c)] \cup [b_{\alpha,\gamma}(c),\infty), & 1 < c < c^*, \\ [0,\infty), & c \geqslant c^*. \end{cases}$$

In the case $1 < c < c^*$, one has $0 < a_{\alpha,\gamma}(c) < b_{\alpha,\gamma}(c)$.

REMARK 3.1. In the cases of Lemma 3.1 where the real roots $a_{\alpha,\gamma}(c)$ and $b_{\alpha,\gamma}(c)$ exist, they can be calculated explicitly using branches W_0 and W_{-1} of the Lambert function:

(3.4)
$$\begin{aligned} a_{\alpha,\gamma}(c) &= -1 - \frac{\alpha}{\gamma} W_0\left(-\frac{\gamma c^{1/\alpha} e^{-\gamma/\alpha}}{\alpha}\right), \\ b_{\alpha,\gamma}(c) &= -1 - \frac{\alpha}{\gamma} W_{-1}\left(-\frac{\gamma c^{1/\alpha} e^{-\gamma/\alpha}}{\alpha}\right). \end{aligned}$$

Next our goal is to apply Lemma 3.1 to the sets $A_\pm$ defined in (3.3).

To this end, we need to consider the following function:

$$c_{\mathrm{lower}}(n) = \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)} \left(\frac{2}{\nu}\right)^{n/2} d_-^{n/2}, \tag{3.5}$$

which has a simpler behavior than c defined in (3.2), and moreover $c(n) \geqslant c_{\mathrm{lower}}(n)$ for all $n \in \mathbb{N}$. Throughout this section, we write c and α in place of $c(n)$ and $\alpha(n)$, with the dependence on n understood. The following three lemmas study $c_{\mathrm{lower}}(n)$; together they verify the hypotheses ($c \geqslant 1$, $\alpha > \gamma_\pm$) that make Lemma 3.1 yield the two-interval form of $A_\pm$ for $n \geqslant n_0$.

LEMMA 3.2. *If $c_{\mathrm{lower}}(1) \geqslant 1$, then c_{lower} is strictly increasing as a function on $\mathbb{N}$.*

LEMMA 3.3. *Let $n_0 = 2k$, where k satisfies*

$$\frac{1}{k} \sum_{l=0}^{k-1} \ln\left(1 + \frac{2l}{\nu}\right) \geqslant -\ln d_-. \tag{3.6}$$

Then $c_{\mathrm{lower}}(n_0) \geqslant 1$.

REMARK 3.2. Note that

$$k^{-1} \sum_{l=0}^{k-1} \ln(1 + 2l/\nu) \sim \ln k \to \infty$$

as $k \to \infty$. Therefore, such an n_0 always exists.

Henceforth, n_0 denotes the smallest integer satisfying the conditions of Lemma 3.3.

In the next lemma we study the monotonicity of the function c_{lower} on $[0, \infty)$.

LEMMA 3.4.

(1) *If $d_- \geqslant 1$, then $c_{\mathrm{lower}}(n)$ is strictly increasing as a function on $\mathbb{N}$, and moreover $c_{\mathrm{lower}}(n) \geqslant 1$ for all $n \geqslant 2$.*

(2) *If $d_- < 1$, then the function $c_{\mathrm{lower}}(x)$ has one minimum point $x_{\min} > \nu/d_- - \nu$ in $[0, \infty)$, and moreover $c_{\mathrm{lower}}(x)$ strictly decreases for $x < x_{\min}$ and strictly increases for $x > x_{\min}$.*

Note that if $d_- \geqslant 1$, then clearly $n_0 = 2$. The next result discusses the case $d_- < 1$.

COROLLARY 3.1. *If $d_- < 1$, then $n_0 > x_{\min}$, where $x_{\min}$ is as in Lemma 3.4.*

The final lemma ensures $c < c^*(\gamma_-)$, so that A_- is a genuine two-interval set rather than all of $[0, \infty)$.

LEMMA 3.5. *$c < e^{\gamma_-}(\alpha/\gamma_- e)^\alpha$ for all $n \geqslant 1$.*

REMARK 3.3. It turns out that without any additional information on d_i both cases $c \geqslant e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha$ and $c < e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha$ are possible even for large n. Indeed, one can check that

$$\frac{c}{e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha} \sim Cn^{-1/2} \prod_{i=1}^{n} \left(\frac{d_i}{d_-}\right)^{1/2}$$

for some positive constant C as $n \to \infty$. Thus, for large n, the ratio exceeds 1 for $d_1 = 1, d_2 = \cdots = d_n = 2$ but stays below 1 for $d_1 = \cdots = d_n = d_- = d_+$.

Now we describe the sets $D_{\alpha,\gamma_\pm}(c)$ using Lemmas 3.1–3.5 and Corollary 3.1.

COROLLARY 3.2. *Under the conditions above for all $n \geqslant n_0$ and $(d_i, i \geqslant 1)$, we have $D_{\alpha,\gamma_-}(c) = [0, a_{\alpha,\gamma_-}(c)] \cup [b_{\alpha,\gamma_-}(c), \infty)$ with $0 \leqslant a_{\alpha,\gamma_-}(c) < b_{\alpha,\gamma_-}(c)$. Moreover, if there exists $\mathrm{n}_{\mathrm{upper}} > 0$ such that*

$$c < e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha \quad \text{for } n \geqslant \mathrm{n}_{\mathrm{upper}}, \tag{3.7}$$

then for all $n \geqslant \max(n_0, \mathrm{n}_{\mathrm{upper}})$, we have $D_{\alpha,\gamma_+}(c) = [0, a_{\alpha,\gamma_+}(c)] \cup [b_{\alpha,\gamma_+}(c), \infty)$ with $0 \leqslant a_{\alpha,\gamma_+}(c) < b_{\alpha,\gamma_+}(c)$. If instead there exists n_{lower} such that

$$c \geqslant e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha \quad \text{for } n \geqslant n_{\mathrm{lower}}, \tag{3.8}$$

then $D_{\alpha,\gamma_+}(c) = [0, \infty)$ for all $n \geqslant \max(n_{\mathrm{lower}}, n_0)$.

From Corollary 3.2, it follows that

$$A_- = \{t \in [0,\infty) \colon t \leqslant \nu a_{\alpha,\gamma_-}(c)\} \cup \{t \in [0,\infty) \colon t \geqslant \nu b_{\alpha,\gamma_-}(c)\}.$$

Under condition (3.7), we also have

$$A_+ = \{t \in [0,\infty) \colon t \leqslant \nu a_{\alpha,\gamma_+}(c)\} \cup \{t \in [0,\infty) \colon t \geqslant \nu b_{\alpha,\gamma_+}(c)\}.$$

Under condition (3.8), $A_+ = [0, \infty)$.

Before stating bounds, we present an exact formula for the TVD.

THEOREM 3.1 (Exact formula). *Let $\mathcal{E}(g_1, I)$ and $\mathcal{E}(g_2, D)$ be the multivariate centered t-distribution with ν degrees of freedom and $\Sigma = I$, and the multivariate centered normal distribution with $\Sigma = D = \operatorname{diag}(d_1, \ldots, d_n)$, where $0 < d_- \leqslant d_i \leqslant d_+$ for all i. Let σ denote the uniform probability measure on $\mathbb{S}^{n-1}$ and define*

$$\gamma_1(\theta) = \frac{\nu}{2} \sum_{i=1}^{n} \frac{\theta_i^2}{d_i} \quad \text{and} \quad \gamma_2(\theta) = \frac{\nu}{2\sum_{i=1}^{n} \theta_i^2 d_i}.$$

Assume $n \geqslant \max(n_0, \mathrm{n}_{\mathrm{upper}})$ *and condition* (3.7) *holds. Then*

$$D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) = \int_{\mathbb{S}^{n-1}} \left[P\left(\gamma_2 b_{\alpha,\gamma_2}; \frac{n}{2}\right) - P\left(\gamma_2 a_{\alpha,\gamma_2}; \frac{n}{2}\right) + I\left(\frac{1}{1+b_{\alpha,\gamma_1}}; \frac{\nu}{2}, \frac{n}{2}\right) - I\left(\frac{1}{1+a_{\alpha,\gamma_1}}; \frac{\nu}{2}, \frac{n}{2}\right) \right] d\sigma(\theta),$$

where $a_{\alpha,\gamma}$ *and* $b_{\alpha,\gamma}$ *are given by* (3.4).

REMARK 3.4. The exact formula requires integration over the unit sphere $\mathbb{S}^{n-1}$, which becomes computationally expensive as n grows. The bounds in Theorem 3.2 below avoid this by optimizing $\gamma_1(\theta)$ and $\gamma_2(\theta)$ over θ, using that $\gamma_1 \in [\nu/(2d_+), \nu/(2d_-)]$ and $\gamma_2 \in [\nu/(2d_+), \nu/(2d_-)]$.

Combining the description of $A_\pm$ with Proposition 2.1, we obtain bounds for the total variation distance.

Figure 1 illustrates these results for $\nu = 5$ and six sequences $d = (d_i)_{i \geqslant 1}$ fluctuating around 1 at different rates. The bounds (dashed lines) come from the closed-form expressions of Theorem 3.2 with $d_- = \min_{i \leqslant n} d_i$, $d_+ = \max_{i \leqslant n} d_i$, shown for $n \geqslant \max(n_0, \mathrm{n}_{\mathrm{upper}})$ under condition (3.7). The Monte Carlo estimates (gray squares) use $D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) = \frac{1}{2}\mathbb{E}\{|1 - q(\boldsymbol{X})/p(\boldsymbol{X})|\}$, $\boldsymbol{X} \sim \mathcal{E}(g_1, I)$, with p, q the densities of $\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)$, approximated by the empirical mean over 5×10^4 draws $\boldsymbol{X} = \boldsymbol{Z}/\sqrt{V/\nu}$, $\boldsymbol{Z} \sim \mathcal{N}(0, I_n)$, $V \sim \chi^2_\nu$. The exact values (green crosses) approximate the $\mathbb{S}^{n-1}$-integral in Theorem 3.1 over 5×10^3 uniform points. Both lie within the bounds in every panel.

THEOREM 3.2. *Let* $\mathcal{E}(g_1, I)$ *and* $\mathcal{E}(g_2, D)$ *be the multivariate centered t-distribution and the multivariate centered normal distribution defined by their densities* (3.1).

(1) *If* (3.8) *is satisfied, then for* $n \geqslant \max(n_0, n_{\mathrm{lower}})$,

$$D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) \leqslant P\left(\frac{\nu b_{\alpha,\gamma_-}}{2d_+}; \frac{n}{2}\right) - P\left(\frac{\nu a_{\alpha,\gamma_-}}{2d_+}; \frac{n}{2}\right)$$

(2) *If* (3.7) *is satisfied, then for* $n \geqslant \max(n_0, \mathrm{n}_{\mathrm{upper}})$, *denoting*

$$C_\pm = P\left(\frac{\nu b_{\alpha,\gamma_\mp}}{2d_\pm}; \frac{n}{2}\right) - P\left(\frac{\nu a_{\alpha,\gamma_\mp}}{2d_\pm}; \frac{n}{2}\right) + I\left(\frac{1}{1+b_{\alpha,\gamma_\pm}}; \frac{\nu}{2}, \frac{n}{2}\right) - I\left(\frac{1}{1+a_{\alpha,\gamma_\pm}}; \frac{\nu}{2}, \frac{n}{2}\right),$$

we have $C_- \leqslant D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) \leqslant C_+$, *where* P *is the lower regularized incomplete Gamma function and* I *is the regularized incomplete Beta function.*

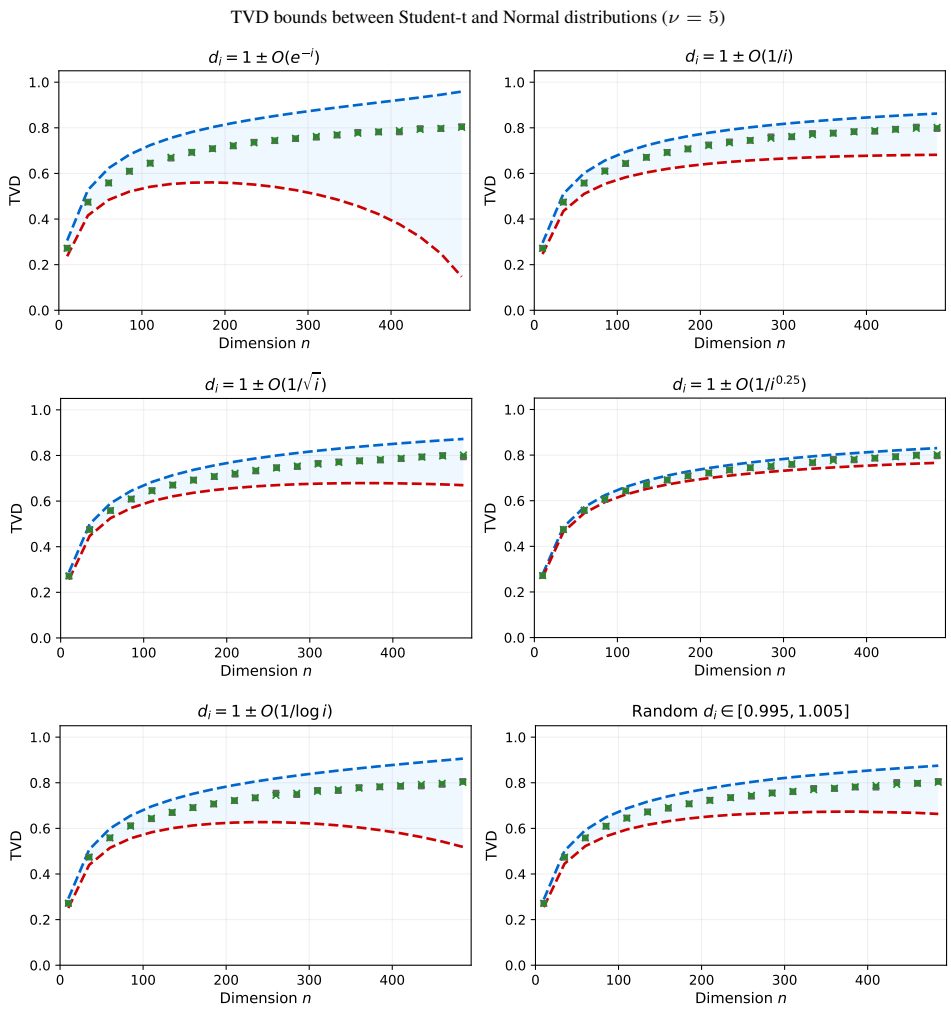


Figure 1. Upper and lower bounds from Theorem 3.2 (dashed lines) compared with Monte Carlo estimates using 5×10^4 samples (gray squares) and the exact formula from Theorem 3.1 (green crosses) for six different sequences $d = (d_i)_{i=1,\dots,n}$ with $\nu = 5$. The shaded region shows the gap between bounds.

REMARK 3.5. Since $x \mapsto I(x; a, b)$ is increasing, the upper bound in (2) of the last theorem is lower than that of (1).

THEOREM 3.3 (Concentration-based bounds). *Let $\mathcal{E}(g_1, I)$ and $\mathcal{E}(g_2, D)$ be the multivariate centered t-distribution and the multivariate centered normal distribution defined by their densities* (3.1)*, where $D = \operatorname{diag}(d_1, \dots, d_n)$ with $0 < d_- \leqslant d_i \leqslant d_+$ for all i. Let $\bar{d} = \frac{1}{n}\sum_{i=1}^n d_i$ and $\overline{d^{-1}} = \frac{1}{n}\sum_{i=1}^n d_i^{-1}$. For any*

$\kappa > 0$, define the effective gamma parameters

$$\gamma_{\mp}^{S} = \frac{\nu}{2(\bar{d} \pm \kappa/\sqrt{n})}, \quad \gamma_{\pm}^{T} = \frac{\nu}{2}\left(\overline{d^{-1}} \pm \frac{\kappa}{\sqrt{n}}\right),$$

and let $\gamma_{\max}^{\text{eff}} = \max(\gamma_{+}^{S}, \gamma_{+}^{T})$. For $n \geqslant \max(n_0, \mathrm{n}_{\text{upper}})$ such that $\bar{d} > \kappa/\sqrt{n}$, $\alpha \geqslant \gamma_{\max}^{\text{eff}}$, and $c < e^{\gamma_{\max}^{\text{eff}}}(\alpha/\gamma_{\max}^{\text{eff}} e)^{\alpha}$, denoting by $F(\theta)$ the integrand from the exact formula, we have

$$\begin{aligned} L^{\text{eff}} - 2\|F - F^{\text{lower}}\|_{\infty} e^{-c_0\kappa^2} &\leqslant D_{\text{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) \\ &\leqslant U^{\text{eff}} + 2\|F - F^{\text{upper}}\|_{\infty} e^{-c_0\kappa^2}, \end{aligned}$$

where U^{eff} and L^{eff} have the same form as C_+ and C_- in Theorem 3.2*, with $\gamma_{\mp}$ replaced by $\gamma_{\mp}^{S}$ and $\gamma_{\pm}$ replaced by $\gamma_{\pm}^{T}$; $F^{\text{upper}}, F^{\text{lower}}$ are defined in the proof; and $c_0 = \frac{n-1}{n}\min(1/(8d_{+}^{2}), d_{-}^{2}/8)$. Bounding $\|F - F^{\text{upper}}\|_{\infty}, \|F - F^{\text{lower}}\|_{\infty} \leqslant 4$ yields the simplified bounds*

$$L^{\text{eff}} - 8e^{-c_0\kappa^2} \leqslant D_{\text{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) \leqslant U^{\text{eff}} + 8e^{-c_0\kappa^2}.$$

The deterministic gap $U^{\text{eff}} - L^{\text{eff}}$ depends on the separation of c from the critical levels $e^{\gamma}(\alpha/\gamma e)^{\alpha}$; the theorem therefore states the concentration bounds in finite-dimensional form.

REMARK 3.6. The parameter κ controls a trade-off. Increasing κ decreases the exceptional-set corrections $e^{-c_0\kappa^2}$, but it also widens the effective intervals for γ_1 and γ_2, and can make the condition $c < e^{\gamma_{\max}^{\text{eff}}}(\alpha/\gamma_{\max}^{\text{eff}} e)^{\alpha}$ fail. Thus κ should be chosen subject to the hypotheses of the theorem. In low-dimensional settings Theorem 3.2 may be preferable.

Proof of Theorem 3.3 (sketch). Define $S(\theta) = \sum_{i=1}^{n} \theta_i^2 d_i$ and $T(\theta) = \sum_{i=1}^{n} \theta_i^2/d_i$. For θ uniform on $\mathbb{S}^{n-1}$, concentration of measure on the sphere implies that $S(\theta)$ and $T(\theta)$ concentrate around their means with deviation $O(1/\sqrt{n})$. Define separate concentration sets G_+ and G_- for upper and lower bounds. On G_+, $\gamma_2(\theta) \geqslant \gamma_{-}^{S}$ and $\gamma_1(\theta) \leqslant \gamma_{+}^{T}$; reverse inequalities hold on G_-. The monotonicity of F_1 and F_2 then yields $F \leqslant F^{\text{upper}}$ on G_+ and $F \geqslant F^{\text{lower}}$ on G_-. The complementary sets have exponentially small measure, contributing error terms bounded by $2\|F - F^{\text{upper/lower}}\|_{\infty} e^{-c_0\kappa^2}$. ∎

3.2. KL divergence. Before stating the next theorem, we introduce auxiliary notation that will be used in its formulation. For $m \in \mathbb{N}$, define

$$\begin{aligned} \alpha_{\mp,m} &= \frac{2d_{\mp}}{\nu} \int_{0}^{\infty} e^{-y}\left(1 + \frac{2d_{\mp}y}{\nu}\right)^{-1} y^{m/2}\, dy \\ &= 2^{-m/2} e^{\frac{\nu}{2d_{\mp}}} \Gamma\left(\frac{m}{2} + 1\right)\left(\frac{d_{\mp}}{\nu}\right)^{-m/2} \Gamma\left(-\frac{m}{2}, \frac{\nu}{2d_{\mp}}\right) \end{aligned} \tag{3.9}$$

and for $n \in \mathbb{N}$ let

$$\Delta_{\mp,n} = \begin{cases} \sqrt{\pi}\left(\pi \operatorname{erfi}\left(\sqrt{\dfrac{\nu}{2d_{\mp}}}\right) + \ln\left(\dfrac{d_{\mp}}{2\nu}\right) - \gamma_{\mathrm{EM}}\right) \\ \qquad\qquad - \dfrac{\sqrt{\pi}\nu}{d_{\mp}} {}_2F_2\left(1,1;\dfrac{3}{2},2;\dfrac{\nu}{2d_{\mp}}\right) & \text{if } n = 1, \\ -e^{\frac{\nu}{2d_{\mp}}} \operatorname{Ei}\left(-\dfrac{\nu}{2d_{\mp}}\right) & \text{if } n = 2, \end{cases} \tag{3.10}$$

and

$$\Delta_{\mp,n} = \begin{cases} \Gamma(n/2)\Delta_{\mp,2} + \Gamma(n/2) \displaystyle\sum_{l=1}^{n/2-1} \frac{\alpha_{\mp,n-2l}}{\Gamma(n/2-l+1)} & \text{if } 2 \mid n,\ n > 2, \\ \dfrac{\Gamma(n/2)}{\sqrt{\pi}}\Delta_{\mp,1} + \Gamma(n/2) \displaystyle\sum_{l=1}^{(n-1)/2} \frac{\alpha_{\mp,n-2l}}{\Gamma(n/2-l+1)} & \text{if } 2 \nmid n,\ n > 1. \end{cases} \tag{3.11}$$

Here γ_{EM} is the Euler–Mascheroni constant.

The next theorem gives the exact forward divergence $D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D))$ (finite if and only if $\nu > 2$) and two-sided bounds for the reverse divergence.

THEOREM 3.4. *Under the assumptions given in this section:*

(1) *If $\nu > 2$, then*

$$\begin{aligned} D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D)) = {} & \frac{1}{2}\sum_{i=1}^{n} \ln d_i + \frac{n}{2}\ln(2\pi) + \frac{\nu}{\nu-2}\sum_{i=1}^{n}\frac{1}{2d_i} \\ & - \frac{1}{2}(n+\nu)\left(\psi_0\left(\frac{n+\nu}{2}\right) - \psi_0\left(\frac{\nu}{2}\right)\right) \\ & + \ln\left(\frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\,\pi^{n/2}\nu^{n/2}}\right). \end{aligned}$$

(2) *If $\nu \leqslant 2$, then $D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D)) = \infty$.*

(3) *For all $\nu > 0$, we have*

$$\begin{aligned} & \max\left(0, \frac{\nu+n}{2\Gamma(n/2)}\Delta_{-,n} - \ln\left(\frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\,\pi^{n/2}\nu^{n/2}}\right) - \frac{n}{2} - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\ln d_i\right) \\ & \quad \leqslant D_{\mathrm{KL}}(\mathcal{E}(g_2, D) \,\|\, \mathcal{E}(g_1, I)) \\ & \quad \leqslant \frac{\nu+n}{2\Gamma(n/2)}\Delta_{+,n} - \ln\left(\frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\,\pi^{n/2}\nu^{n/2}}\right) - \frac{n}{2} - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\ln d_i, \end{aligned}$$

where $\Delta_{\mp,n}$ is given in (3.10).

4. DISTANCES BETWEEN TWO GAMMA DISTRIBUTIONS

In this section we derive bounds on the total variation distance between two multivariate Gamma distributions with independent components defined by their pdfs

$$p_1(x) = \prod_{i=1}^{n} \frac{\lambda_i^{\alpha_i} x_i^{\alpha_i - 1} e^{-\lambda_i x_i}}{\Gamma(\alpha_i)} \quad \text{and} \quad p_2(x) = \prod_{i=1}^{n} \frac{\mu_i^{\beta_i} x_i^{\beta_i - 1} e^{-\mu_i x_i}}{\Gamma(\beta_i)},$$

with $\alpha_i, \beta_i, \lambda_i, \mu_i > 0$ for $i = 1, \dots, n$. Before we formulate our next result, recall the Berry–Esseen inequality.

Let Z_i, $i = 1, 2, \dots$, be a sequence of independent random variables such that

$$\mathbb{E}\{Z_i\} = 0, \quad \mathbb{E}\{Z_i^2\} = \sigma_i^2 \quad \text{and} \quad \mathbb{E}\{|Z_i|^3\} = \rho_i < \infty, \quad i = 1, 2, \dots. \tag{4.1}$$

Then according to the Berry–Esseen theorem there exists an absolute constant C_0 such that

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leqslant C_0 \kappa_n, \tag{4.2}$$

where F_n and Φ are distribution functions of

$$S_n = (Z_1 + \cdots + Z_n) / \sqrt{\sigma_1^2 + \cdots + \sigma_n^2} \quad \text{and} \quad Z \sim N(0, 1)$$

respectively, and $\kappa_n = (\sum_{i=1}^n \rho_i) / (\sum_{i=1}^n \sigma_i^2)^{3/2}$.

Denote by $\mathbb{P}_{\boldsymbol{X}}$ and $\mathbb{P}_{\boldsymbol{Y}}$ the distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with pdfs p_1 and p_2 respectively.

THEOREM 4.1.

$$D_{\mathrm{TV}}(\mathbb{P}_{\boldsymbol{X}}, \mathbb{P}_{\boldsymbol{Y}}) = \mathbb{P}\left\{ \frac{c(n) - \sum_{i=1}^n \mathbb{E}\{Z_i\}}{\sqrt{\sum_{i=1}^n \mathrm{Var}\{Z_i\}}} \leqslant Z \leqslant \frac{c(n) - \sum_{i=1}^n \mathbb{E}\{\widetilde{Z}_i\}}{\sqrt{\sum_{i=1}^n \mathrm{Var}\{\widetilde{Z}_i\}}} \right\} + \varepsilon,$$

where

$$\begin{aligned}
Z_i &= (\alpha_i - \beta_i) \ln X_i + (\mu_i - \lambda_i) X_i, \quad X_i \sim \mathrm{Gamma}(\alpha_i, \lambda_i), \\
\widetilde{Z}_i &= (\alpha_i - \beta_i) \ln Y_i + (\mu_i - \lambda_i) Y_i, \quad Y_i \sim \mathrm{Gamma}(\beta_i, \mu_i), \\
&\qquad\qquad\quad Z \sim N(0, 1), \\
\mathbb{E}\{Z_i\} &= (\alpha_i - \beta_i)(\psi_0(\alpha_i) - \ln \lambda_i) + \alpha_i \left(\frac{\mu_i}{\lambda_i} - 1 \right), \\
\mathrm{Var}\{Z_i\} &= (\alpha_i - \beta_i)^2 \psi_1(\alpha_i) - \frac{(\lambda_i - \mu_i)(\alpha_i(\lambda_i + \mu_i) - 2\beta_i \lambda_i)}{\lambda_i^2}, \\
\mathbb{E}\{\widetilde{Z}_i\} &= (\alpha_i - \beta_i)(\psi_0(\beta_i) - \ln \mu_i) + \beta_i \left(1 - \frac{\lambda_i}{\mu_i} \right),
\end{aligned}$$

$$\operatorname{Var}\{\widetilde{Z}_i\} = (\alpha_i - \beta_i)^2 \psi_1(\beta_i) + \frac{(\lambda_i - \mu_i)(\beta_i(\lambda_i + \mu_i) - 2\alpha_i \mu_i)}{\mu_i^2},$$

$$c(n) = \sum_{i=1}^{n} \ln(\Gamma(\alpha_i)\mu_i^{\beta_i}/(\Gamma(\beta_i)\lambda_i^{\alpha_i})),$$

$$|\varepsilon| \leqslant C_0(\kappa_n + \widetilde{\kappa}_n), \quad \kappa_n = \frac{\sum_{i=1}^n \rho_i}{(\sum_{i=1}^n \operatorname{Var}\{Z_i\})^{3/2}},$$

$$\widetilde{\kappa}_n = \frac{\sum_{i=1}^n \widetilde{\rho}_i}{(\sum_{i=1}^n \operatorname{Var}\{\widetilde{Z}_i\})^{3/2}}$$

and $\rho_i = \mathbb{E}\{|Z_i - \mathbb{E}\{Z_i\}|^3\}$ *and* $\widetilde{\rho}_i = \mathbb{E}\{|\widetilde{Z}_i - \mathbb{E}\{\widetilde{Z}_i\}|^3\}$.

5. APPENDIX – PROOFS

Proof of Lemma 2.1. Diagonalize $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} = P^\top D P$ and apply the change of variables $\boldsymbol{y} = P\Sigma_1^{-1/2}\boldsymbol{x}$. Under this substitution, $\boldsymbol{x}^\top\Sigma_1^{-1}\boldsymbol{x} = |\boldsymbol{y}|^2$ and $\boldsymbol{x}^\top\Sigma_2^{-1}\boldsymbol{x} = \boldsymbol{y}^\top D^{-1}\boldsymbol{y}$, while $d\boldsymbol{x} = \sqrt{\det\Sigma_1}\, d\boldsymbol{y}$. The Jacobian and determinant factors simplify to give

$$D_f(\mathcal{E}(g_1, \Sigma_1) \,\|\, \mathcal{E}(g_2, \Sigma_2)) = D_f(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D)).$$

The eigenvalues of $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}$ and $\Sigma_2\Sigma_1^{-1}$ coincide, so D is the eigenvalue matrix of $\Sigma_1^{-1}\Sigma_2$. ∎

Proof of Proposition 2.1. We start with the properties of total variation distance.

(a) Follows from [7, Theorem 2.9].

(b) Using (2.1), the TVD equals

$$\mathbb{P}\{p_1(\boldsymbol{X}) \geqslant p_2(\boldsymbol{X})\} - \mathbb{P}\{p_1(\boldsymbol{Y}) \geqslant p_2(\boldsymbol{Y})\}.$$

Since $|\boldsymbol{X}|^2/d_+ \leqslant \sum X_i^2/d_i \leqslant |\boldsymbol{X}|^2/d_-$, and g_2 is a decreasing function on $[0, \infty)$, we have, for any $\boldsymbol{x} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathbb{P}\{|\boldsymbol{X}|^2 \in A_-\} &= \mathbb{P}\Big\{\prod_{i=1}^n d_i^{1/2} g_1(|\boldsymbol{X}|^2) \geqslant g_2(|\boldsymbol{X}|^2/d_+)\Big\} \\
&\leqslant \mathbb{P}\Big\{\prod_{i=1}^n d_i^{1/2} g_1(|\boldsymbol{X}|^2) \geqslant g_2\Big(\sum X_i^2/d_i\Big)\Big\} \\
&\leqslant \mathbb{P}\Big\{\prod_{i=1}^n d_i^{1/2} g_1(|\boldsymbol{X}|^2) \geqslant g_2(|\boldsymbol{X}|^2/d_-)\Big\} \\
&\leqslant \mathbb{P}\{|\boldsymbol{X}|^2 \in A_+\}
\end{aligned}$$

Similarly for $\boldsymbol{Y}$: since g_1 is decreasing and $d_- \sum Y_i^2/d_i \leqslant |\boldsymbol{Y}|^2 \leqslant d_+ \sum Y_i^2/d_i$, we have

$$\begin{aligned}\mathbb{P}\Big\{d_+\textstyle\sum Y_i^2/d_i \in A_-\Big\} &= \mathbb{P}\Big\{\prod_{i=1}^n d_i^{1/2} g_1\Big(d_+\textstyle\sum Y_i^2/d_i\Big) \geqslant g_2\Big(\textstyle\sum Y_i^2/d_i\Big)\Big\}\\ &\leqslant \mathbb{P}\Big\{\prod_{i=1}^n d_i^{1/2} g_1(|\boldsymbol{Y}|^2) \geqslant g_2\Big(\textstyle\sum Y_i^2/d_i\Big)\Big\}\\ &\leqslant \mathbb{P}\Big\{\prod_{i=1}^n d_i^{1/2} g_1\Big(d_-\textstyle\sum Y_i^2/d_i\Big) \geqslant g_2\Big(\textstyle\sum Y_i^2/d_i\Big)\Big\}.\\ &\leqslant \mathbb{P}\Big\{d_-\textstyle\sum Y_i^2/d_i \in A_+\Big\}.\end{aligned}$$

Combining the above yields the two-sided bound. Converting to spherical coordinates we get

$$\mathbb{P}\{|\boldsymbol{X}|^2 \in A_\pm\} = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \mathbf{1}\{r^2 \in A_\pm\} g_1(r^2)\, r^{n-1}\, dr.$$

For the $\boldsymbol{Y}$ terms, set $\boldsymbol{z} = D^{-1/2}\boldsymbol{y}$, or equivalently $\boldsymbol{y} = D^{1/2}\boldsymbol{z}$. Then $d\boldsymbol{y} = \sqrt{\det D}\, d\boldsymbol{z}$, which cancels the factor $1/\sqrt{\det D}$ in the density of $\mathcal{E}(g_2, D)$, and $\sum y_i^2/d_i$ becomes $|\boldsymbol{z}|^2$. Hence

$$\mathbb{P}\Big\{d_\mp \textstyle\sum Y_i^2/d_i \in A_\pm\Big\} = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \mathbf{1}\{d_\mp r^2 \in A_\pm\} g_2(r^2) r^{n-1}\, dr.$$

Now we check the inequalities for KL-divergence. Part (a) follows by direct calculation using spherical coordinates.

For part (b), by Lemma 1 we may assume $\Sigma_1 = I$ and $\Sigma_2 = D$. Using the fact that g_2 is decreasing on $[0, \infty)$, we have, for any $\boldsymbol{x} \in \mathbb{R}^n$,

$$g_2\left(\frac{|\boldsymbol{x}|^2}{d_+}\right) \geqslant g_2\left(\sum_{i=1}^n \frac{x_i^2}{d_i}\right) \geqslant g_2\left(\frac{|\boldsymbol{x}|^2}{d_-}\right).$$

Now observe that

$$D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D)) = \int_{\mathbb{R}^n} g_1(|\boldsymbol{x}|^2) \ln\left(\frac{g_1(|\boldsymbol{x}|^2) \prod_{i=1}^n d_i^{1/2}}{g_2(\sum_{i=1}^n x_i^2/d_i)}\right) d\boldsymbol{x}.$$

Using the bounds on g_2 and converting to spherical coordinates yields the result. ∎

Proof of Lemma 3.1. Let $h(x) = \log \varphi_{\alpha,\gamma}(x) = \alpha \log(1+x) - \gamma x$. Then $h'(x) = \alpha/(1+x) - \gamma$ and $h''(x) = -\alpha/(1+x)^2 < 0$. If $\alpha \leqslant \gamma$, then h is decreasing on $[0, \infty)$, so $\varphi_{\alpha,\gamma}(0) = 1$ is the maximum and the asserted description of $D_{\alpha,\gamma}(c)$ follows.

If $\alpha > \gamma$, then h is increasing on $[0, (\alpha-\gamma)/\gamma]$ and decreasing afterwards. Hence $\varphi_{\alpha,\gamma}$ attains its maximum $c^* = e^\gamma(\alpha/\gamma e)^\alpha > 1$ at $x_0 = (\alpha-\gamma)/\gamma$, while $\varphi_{\alpha,\gamma}(0) = 1$ and $\varphi_{\alpha,\gamma}(x) \to 0$ as $x \to \infty$. This gives the stated sublevel-set cases, except for the explicit root formulas.

Solving $\varphi_{\alpha,\gamma}(x) = c$ with $y = -(\gamma/\alpha)(1+x)$ gives

$$ye^{y} = -\frac{\gamma c^{1/\alpha} e^{-\gamma/\alpha}}{\alpha} =: z.$$

The condition $z \in [-1/e, 0]$ is equivalent to $c \leqslant e^{\gamma}(\alpha/\gamma e)^{\alpha}$. Therefore, in the real-root regimes, the two roots are given by the Lambert branches $W_0(z)$ and $W_{-1}(z)$, yielding (3.4). Since $W_{-1}(z) \leqslant W_0(z)$, we have $a_{\alpha,\gamma}(c) \leqslant b_{\alpha,\gamma}(c)$.

It remains to determine the sign of $a_{\alpha,\gamma}(c)$. Put $q = \gamma/\alpha$. If $0 < c < 1$ and $q \leqslant 1$, then $z > -qe^{-q}$, hence by monotonicity of W_0, $W_0(z) > -q$ and $a_{\alpha,\gamma}(c) = -1 - q^{-1}W_0(z) < 0$. If $0 < c < 1$ and $q > 1$, then $W_0(z) \geqslant -1$, so $a_{\alpha,\gamma}(c) \leqslant -1 + q^{-1} < 0$. Thus $D_{\alpha,\gamma}(c) = [b_{\alpha,\gamma}(c), \infty)$ for $0 < c < 1$.

Finally, if $\alpha > \gamma$ and $1 < c < c^{*}$, then $q < 1$ and $z < -qe^{-q}$, while $z > -1/e$. Therefore $W_0(z) < -q$, and $a_{\alpha,\gamma}(c) = -1 - q^{-1}W_0(z) > 0$. At $c = 1$, the same formula gives $a_{\alpha,\gamma}(1) = 0$. ∎

Proof of Lemma 3.2. It suffices to show $c_{\mathrm{lower}}(2k+1)/c_{\mathrm{lower}}(2k) > 1$ and $c_{\mathrm{lower}}(2k)/c_{\mathrm{lower}}(2k-1) > 1$ for all $k \in \mathbb{N}$. Using the Gamma function identity

$$\Gamma\!\left(\frac{\nu+1}{2}+k\right)\!/\Gamma\!\left(\frac{\nu}{2}+k\right) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \prod_{l=0}^{k-1}\left(1+\frac{1}{\nu+2l}\right),$$

one obtains

$$c_{\mathrm{lower}}(2k+1)/c_{\mathrm{lower}}(2k) = c_{\mathrm{lower}}(1) \prod_{l=0}^{k-1}\left(1+\frac{1}{\nu+2l}\right) \geqslant c_{\mathrm{lower}}(1) \geqslant 1.$$

For the even-to-odd ratio, logarithmic convexity of Γ ($\Gamma(x)\Gamma(x+1) \geqslant \Gamma(x+1/2)^2$) gives

$$c_{\mathrm{lower}}(2k)/c_{\mathrm{lower}}(2k-1) \geqslant c_{\mathrm{lower}}(1) \prod_{l=0}^{k-2}\left(1+\frac{1}{\nu+2l+1}\right) > 1. \quad ∎$$

Proof of Lemma 3.3. For $n_0 = 2k$, Gamma identities give

$$c_{\mathrm{lower}}(n_0) = \prod_{l=0}^{k-1}(1+2l/\nu)\cdot d_{-}^{k}.$$

Taking logarithms, we see that condition (3.6) yields $\ln c_{\mathrm{lower}}(n_0) \geqslant 0$. ∎

Proof of Lemma 3.4. We have

$$c'_{\mathrm{lower}}(x) = \widetilde{f}(x)\big(\ln(2d_{-}/\nu) + \psi_0((\nu+x)/2)\big)$$

with

$$\widetilde{f}(x) = 2^{-1+x/2} d_{-}^{x/2} \Gamma((\nu+x)/2)/(\nu^{x/2}\Gamma(\nu/2)) > 0.$$

Since ψ_0 is strictly increasing, c_{lower} has at most one critical point. If $d_{-} < 1$, using $\psi_0(x) < \ln x$ gives $c'_{\mathrm{lower}}(0) < \widetilde{f}(0)\ln d_{-} < 0$, so the minimum point $x_{\min}$

exists. Moreover,

$$0 = c'_{\text{lower}}(x_{\min}) < \widetilde{f}(x_{\min}) \ln(d_-(\nu + x_{\min})/\nu),$$

hence $x_{\min} > \nu/d_- - \nu$. ∎

Proof of Corollary 3.1. Assume for contradiction that $n_0 \leqslant x_{\min}$. Then by Lemma 3.4, the function c_{lower} is strictly decreasing on $[0, n_0]$. Since $n_0 \geqslant 2$, Lemma 3.3 gives

$$c_{\text{lower}}(1) > c_{\text{lower}}(2) \geqslant c_{\text{lower}}(n_0) \geqslant 1.$$

In particular, $c_{\text{lower}}(1) \geqslant 1$. Lemma 3.2 then implies that c_{lower} is strictly increasing on $\mathbb{N}$, so $c_{\text{lower}}(1) < c_{\text{lower}}(2)$, a contradiction. ∎

Proof of Lemma 3.5. Write $c/(e^{\gamma_-}(\alpha/\gamma_- e)^\alpha) = C_2 \widetilde{f}(n)\widetilde{g}(n)$ where $\widetilde{f}(n) = \Gamma((\nu+n)/2)(2e)^{n/2}/(\nu+n)^{(\nu+n)/2}$ and $\widetilde{g}(n) = \prod_{i=1}^n (d_i/d_+)^{1/2}$. Since $\widetilde{g}$ is clearly non-increasing and $\widetilde{f}'(n) \propto \psi_0((\nu+n)/2) - \ln((\nu+n)/2) < 0$, the ratio decreases in n. For $n = 1$, using $\Gamma(x+1/2) \leqslant x^{1/2}\Gamma(x)$ and $(1+1/x)^{x+1} > e$ yields

$$c/(e^{\gamma_-}(\alpha/\gamma_- e)^\alpha) \leqslant (e/(1+1/\nu)^{\nu+1})^{1/2} < 1. \quad ∎$$

Proof of Theorem 3.1. Write $D_{\text{TV}}(p_1, p_2) = Q_1 - Q_2$ where $Q_1 = \mathbb{P}\{p_1(\boldsymbol{X}) \geqslant p_2(\boldsymbol{X})\}$ and $Q_2 = \mathbb{P}\{p_1(\boldsymbol{Y}) \geqslant p_2(\boldsymbol{Y})\}$.

Computation of Q_1. Switching to polar coordinates $x = r\theta$ with $\theta \in \mathbb{S}^{n-1}$ and substituting $t = r^2/\nu$, the indicator $\mathbf{1}\{p_1(x) \geqslant p_2(x)\}$ becomes $\mathbf{1}\{\varphi_{\alpha,\gamma_1(\theta)}(t) \leqslant c\}$ where $\gamma_1(\theta) = \frac{\nu}{2}\sum \theta_i^2/d_i$. By the sublevel set characterization, the complement of this set is $(a_{\alpha,\gamma_1}, b_{\alpha,\gamma_1})$. Integrating the radial part against $t^{n/2-1}/(1+t)^{(\nu+n)/2}$ and using $|\mathbb{S}^{n-1}| = 2\pi^{n/2}/\Gamma(n/2)$, we obtain

$$\begin{aligned} Q_1 = {} & \frac{\Gamma((\nu+n)/2)}{2\Gamma(\nu/2)\pi^{n/2}} B\left(\frac{n}{2}, \frac{\nu}{2}\right) |\mathbb{S}^{n-1}| \\ & - \frac{\Gamma((\nu+n)/2)}{2\Gamma(\nu/2)\pi^{n/2}} \int_{\mathbb{S}^{n-1}} \int_{a_{\alpha,\gamma_1}}^{b_{\alpha,\gamma_1}} \frac{t^{n/2-1}}{(1+t)^{(n+\nu)/2}}\, dt\, dS(\theta). \end{aligned}$$

Computation of Q_2. Analogously, substituting $x_i \rightsquigarrow x_i\sqrt{2d_i}$, switching to polar coordinates, and setting $t = \frac{2r^2}{\nu}\sum \theta_i^2 d_i$ with $\gamma_2(\theta) = \nu/(2\sum \theta_i^2 d_i)$, we get

$$Q_2 = \frac{|\mathbb{S}^{n-1}|}{2\pi^{n/2}}\Gamma\left(\frac{n}{2}\right) - \frac{1}{2\pi^{n/2}} \int_{\mathbb{S}^{n-1}} \int_{\gamma_2 a_{\alpha,\gamma_2}}^{\gamma_2 b_{\alpha,\gamma_2}} t^{n/2-1} e^{-t}\, dt\, dS(\theta).$$

Final formula. The constant terms cancel (using $\frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\Gamma(n/2)} B(n/2, \nu/2) = 1$). Normalizing by $d\sigma = dS/|\mathbb{S}^{n-1}|$ and expressing the remaining integrals via regularized incomplete Gamma (P) and Beta (I) functions yields the stated formula. ∎

Proof of Theorem 3.2.

CASE 1: $c \geqslant e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha$. Then $A_+ = [0,\infty)$ and the lower bound from Proposition 2.1 is trivial. For the upper bound, using the fact that p_1 integrates to 1 and $A_-^c = (\nu a_{\alpha,\gamma_-}(c),\, \nu b_{\alpha,\gamma_-}(c))$, we get

$$D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) \leqslant \frac{1}{\Gamma(n/2)} \int\limits_{\nu a_{\alpha,\gamma_-}/(2d_+)}^{\nu b_{\alpha,\gamma_-}/(2d_+)} y^{n/2-1} e^{-y}\, dy$$
$$= P\biggl(\frac{\nu b_{\alpha,\gamma_-}}{2d_+}; \frac{n}{2}\biggr) - P\biggl(\frac{\nu a_{\alpha,\gamma_-}}{2d_+}; \frac{n}{2}\biggr).$$

CASE 2: $c < e^{\gamma_+}(\alpha/\gamma_+ e)^\alpha$. Then both $A_\pm$ have two-component sublevel sets with $A_\pm^c = (\nu a_{\alpha,\gamma_\pm}(c),\, \nu b_{\alpha,\gamma_\pm}(c))$. Applying Proposition 2.1 and rewriting in terms of complements (using that p_1 and p_2 are pdfs), the upper and lower bounds become

$$C_\pm = P\biggl(\frac{\nu b_{\alpha,\gamma_\mp}}{2d_\pm}; \frac{n}{2}\biggr) - P\biggl(\frac{\nu a_{\alpha,\gamma_\mp}}{2d_\pm}; \frac{n}{2}\biggr) - \frac{2\pi^{n/2}}{\Gamma(n/2)} \int\limits_{(\nu a_{\alpha,\gamma_\pm})^{1/2}}^{(\nu b_{\alpha,\gamma_\pm})^{1/2}} g_1(r^2)\, r^{n-1}\, dr.$$

For the integral of g_1, substituting $t = r^2/\nu$ and using the representation $B(x;a,b) = \int_{(1-x)/x}^\infty t^{b-1}(1+t)^{-(a+b)}\, dt$, we obtain

$$\frac{2\pi^{n/2}}{\Gamma(n/2)} \int\limits_{(\nu a_{\alpha,\gamma_\pm})^{1/2}}^{(\nu b_{\alpha,\gamma_\pm})^{1/2}} g_1(r^2)\, r^{n-1}\, dr = I\biggl(\frac{1}{1+a_{\alpha,\gamma_\pm}}; \frac{\nu}{2}, \frac{n}{2}\biggr) - I\biggl(\frac{1}{1+b_{\alpha,\gamma_\pm}}; \frac{\nu}{2}, \frac{n}{2}\biggr),$$

which gives the stated formulas for $C_\pm$. ∎

Proof of Theorem 3.3. By Theorem 3.1,

$$D_{\mathrm{TV}}(\mathcal{E}(g_1, I), \mathcal{E}(g_2, D)) = \int\limits_{\mathbb{S}^{n-1}} F(\theta)\, d\sigma(\theta),$$

where $F(\theta)$ depends on $\gamma_1(\theta) = \frac{\nu}{2}\sum \theta_i^2/d_i$ and $\gamma_2(\theta) = \nu/(2\sum \theta_i^2 d_i)$, and $|F| \leqslant 2$.

Concentration. Define $S(\theta) = \sum \theta_i^2 d_i$ and $T(\theta) = \sum \theta_i^2/d_i$, with Lipschitz constants $2d_+$ and $2d_-^{-1}$ respectively. By concentration of measure on $\mathbb{S}^{n-1}$ [14], the sets

$$G_+ = \{S \leqslant \bar{d} + \kappa/\sqrt{n}, T \leqslant \overline{d^{-1}} + \kappa/\sqrt{n}\},$$
$$G_- = \{S \geqslant \bar{d} - \kappa/\sqrt{n},\, T \geqslant \overline{d^{-1}} - \kappa/\sqrt{n}\}$$

satisfy $\sigma(\mathbb{S}^{n-1} \setminus G_\pm) \leqslant 2e^{-c_0\kappa^2}$ with $c_0 = \frac{n-1}{n}\min(1/(8d_+^2),\, d_-^2/8)$.

Monotonicity and decomposition. Write $F = F_1 - F_2$, where

$$F_1(\theta) = P\biggl(\gamma_2 b_{\alpha,\gamma_2}; \frac{n}{2}\biggr) - P\biggl(\gamma_2 a_{\alpha,\gamma_2}; \frac{n}{2}\biggr),$$

$$F_2(\theta) = I\Big(\frac{1}{1+a_{\alpha,\gamma_1}}; \frac{\nu}{2}, \frac{n}{2}\Big) - I\Big(\frac{1}{1+b_{\alpha,\gamma_1}}; \frac{\nu}{2}, \frac{n}{2}\Big).$$

Since $(1+x)^\alpha e^{-\gamma x}$ is decreasing in γ for $x > 0$, the interval $[a_{\alpha,\gamma}, b_{\alpha,\gamma}]$ shrinks as γ increases, so F_1 is decreasing in γ_2 and F_2 is decreasing in γ_1, hence F is monotone in γ_1, γ_2. Define constant bounds

$$F^{\text{upper}} = F_1|_{\gamma_2=\gamma_-^S} - F_2|_{\gamma_1=\gamma_+^T} \quad \text{and} \quad F^{\text{lower}} = F_1|_{\gamma_2=\gamma_+^S} - F_2|_{\gamma_1=\gamma_-^T}.$$

On G_+, $\gamma_2 \geqslant \gamma_-^S$ and $\gamma_1 \leqslant \gamma_+^T$, so $F_1 \leqslant F_1|_{\gamma_-^S}$ and $F_2 \geqslant F_2|_{\gamma_+^T}$, hence $F \leqslant F^{\text{upper}}$. Analogously, $F \geqslant F^{\text{lower}}$ on G_-. Therefore

$$\begin{aligned}\int F\,d\sigma &\leqslant \int F^{\text{upper}}\,d\sigma + \|F - F^{\text{upper}}\|_\infty\,\sigma(\mathbb{S}^{n-1} \setminus G_+)\\ &\leqslant U^{\text{eff}} + 2\|F - F^{\text{upper}}\|_\infty\, e^{-c_0\kappa^2},\end{aligned}$$

where $U^{\text{eff}} = F^{\text{upper}}$. The lower bound follows analogously. Bounding $\|F - F^{\text{upper/lower}}\|_\infty \leqslant 4$ yields the simplified bounds. ∎

Proof of Theorem 3.4.

PART 1 ($\nu > 2$). Using $g_2(t) = (2\pi)^{-n/2}e^{-t/2}$ and $\mathbb{E}\{X_1^2\} = \nu/(\nu-2)$ shows that

$$\begin{aligned}&D_{\mathrm{KL}}(\mathcal{E}(g_1, I) \,\|\, \mathcal{E}(g_2, D))\\ &= \frac{1}{2}\sum \ln d_i + \frac{n}{2}\ln(2\pi) + \frac{\nu}{\nu-2}\sum \frac{1}{2d_i} + \frac{2\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty g_1(r^2)\ln g_1(r^2)\, r^{n-1}\,dr.\end{aligned}$$

The entropy integral evaluates to

$$-\frac{\nu+n}{2}\big(\psi_0((\nu+n)/2) - \psi_0(\nu/2)\big) + \ln\big(\Gamma((\nu+n)/2)/(\Gamma(\nu/2)\pi^{n/2}\nu^{n/2})\big),$$

yielding the stated formula.

PART 2 (bounds on reverse KL). By Proposition 2.1, the bounds involve $\int_0^\infty e^{-r^2/2}\ln(g_1(d_\mp r^2))\, r^{n-1}\,dr$. Substituting $y = r^2/2$ reduces the task to computing

$$\widetilde{\Delta}_{\mp,n} = \int_0^\infty e^{-y}\ln(1 + 2d_\mp y/\nu)\, y^{n/2-1}\,dy.$$

We show $\widetilde{\Delta}_{\mp,n} = \Delta_{\mp,n}$ by induction on n.

Base cases. For $n = 2$, $\widetilde{\Delta}_{\mp,2} = -e^{\nu/(2d_\mp)}\,\mathrm{Ei}(-\nu/(2d_\mp))$, and for $n = 1$ the integral evaluates to the stated expression involving erfi and ${}_2F_2$; both may be verified directly.

Induction step. Let $n \geqslant 3$ and assume that we know that $\widetilde{\Delta}_{\mp,m} = \Delta_{\mp,m}$ for all $m < n$. We prove the equality for $m = n$. Integration by parts gives the recurrence

$$\widetilde{\Delta}_{\mp,n} = (n/2-1)\widetilde{\Delta}_{\mp,n-2} + \alpha_{\mp,n-2},$$

where the last term is $\alpha_{\mp,n-2}$ by (3.9). Iterating the recurrence k times, for any integer k with $0 < k < n/2$, yields

$$\begin{aligned}\widetilde{\Delta}_{\mp,n} &= \frac{\Gamma(n/2)}{\Gamma(n/2-k)}\widetilde{\Delta}_{\mp,n-2k} + \Gamma(n/2)\sum_{l=1}^{k}\frac{\alpha_{\mp,n-2l}}{\Gamma(n/2-l+1)} \\ &= \frac{\Gamma(n/2)}{\Gamma(n/2-k)}\Delta_{\mp,n-2k} + \Gamma(n/2)\sum_{l=1}^{k}\frac{\alpha_{\mp,n-2l}}{\Gamma(n/2-l+1)},\end{aligned}$$

where the last equality uses the induction hypothesis.

Taking $k = n/2-1$ for even n, or $k = (n-1)/2$ for odd n, reduces the remaining term to the base case 2 or 1, respectively. The definitions in (3.10) therefore give $\widetilde{\Delta}_{\mp,n} = \Delta_{\mp,n}$. ∎

Proof of Theorem 4.1. Taking logarithms of the likelihood ratio, we find that $\{p_1 \geqslant p_2\} = \{\sum_{i=1}^{n} Z_i \geqslant c(n)\}$ where $c(n) = \sum_{i=1}^{n} \ln(\Gamma(\alpha_i)\mu_i^{\beta_i}/(\Gamma(\beta_i)\lambda_i^{\alpha_i}))$. Hence

$$D_{\mathrm{TV}}(\mathbb{P}_{\boldsymbol{X}}, \mathbb{P}_{\boldsymbol{Y}}) = \mathbb{P}\Big\{\sum Z_i \geqslant c(n)\Big\} - \mathbb{P}\Big\{\sum \widetilde{Z}_i \geqslant c(n)\Big\}.$$

Standardizing both sums and applying the Berry–Esseen theorem (4.2) yields the stated formula with $|\varepsilon| \leqslant C_0(\kappa_n + \widetilde{\kappa}_n)$. The formulas for $\mathbb{E}\{Z_i\}$, $\mathrm{Var}\{Z_i\}$, etc. follow from standard Gamma distribution moments. ∎

REFERENCES

[1] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publ., New York, 1992 (reprint of the 1972 edition).

[2] L. Barabesi and L. Pratelli, *An inequality for the total variation distance between high-dimensional centered Gaussian laws*, Statist. Probab. Lett. 211 (2024), art. 110148, 5 pp.

[3] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, *Variational inference: a review for statisticians*, J. Amer. Statist. Assoc. 112 (2017), 859–877.

[4] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, *On the Lambert W function*, Adv. Comput. Math. 5 (1996), 329–359.

[5] G. P. Dehaene, *A deterministic and computable Bernstein–von Mises theorem*, arXiv:1904.02505v2 (2019).

[6] L. Devroye, A. Mehrabian, and T. Reddad, *The total variation distance between high-dimensional Gaussians with the same mean*, arXiv:1810.08693v7 (2023).

[7] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, Monogr. Statist. Appl. Probab. 36, Chapman and Hall, London, 1990.

[8] E. Ghazi and I. Issa, *Total variation meets differential privacy*, IEEE J. Selected Areas Inform. Theory 5 (2024), 207–220.

[9] A. L. Gibbs and F. E. Su, *On choosing and bounding probability metrics*, Internat. Statist. Rev. 70 (2002), 419–435.

[10] A. K. Gupta, T. Varga, and T. Bodnar, *Elliptically Contoured Models in Statistics and Portfolio Theory*, 2nd ed., Springer, New York, 2013.

[11] J. Jia, C. Tan, Z. Liu, X. Li, Z. Liu, S. Lv, and C. Dong, *Total variation distance privacy: Accurately measuring inference attacks and improving utility*, Inform. Sci. 626 (2023), 537–558.

[12] G. Kaissis, J. Hayes, A. Ziller, and D. Rueckert, *Bounding data reconstruction attacks with the hypothesis testing interpretation of differential privacy*, arXiv:2307.03928v1 (2023).

[13] M. J. Kasprzak, R. Giordano, and T. Broderick, *How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences*, J. Mach. Learn. Res. 26 (2025), art. 87.

[14] M. Ledoux, *The Concentration of Measure Phenomenon*, Math. Surveys Monogr. 89, Amer. Math. Soc., Providence, RI, 2001.

[15] J. Li, W. Guo, X. Han, J. Cai, and X. Liu, *Federated learning based on defending against data poisoning attacks in IoT*, arXiv:2209.06397v1 (2022).

[16] DLMF, *NIST Digital Library of Mathematical Functions*, https://dlmf.nist.gov/, Release 1.2.3 of 2024-12-15.

[17] F. Nielsen, *On information projections between multivariate elliptical and location-scale families*, arXiv:2101.03839v2 (2021).

[18] F. Nielsen and R. Nock, *Entropies and cross-entropies of exponential families*, in: 2010 IEEE International Conference on Image Processing, 2010, 3621–3624.

[19] B. Rassouli and D. Gündüz, *Optimal utility-privacy trade-off with total variation distance as a privacy measure*, IEEE Trans. Inform. Forensics Secur. 15 (2020), 594–603.

[20] B. Roig-Solvas and M. Sznaier, *Euclidean distance bounds for linear matrix inequalities analytic centers using a novel bound on the Lambert function*, SIAM J. Control Optim. 60 (2022), 720–731.

[21] I. Sason and S. Verdú, *f-divergence inequalities*, IEEE Trans. Inform. Theory 62 (2016), 5973–6006.

[22] V. Spokoiny, *Dimension free nonasymptotic bounds on the accuracy of high-dimensional Laplace approximation*, SIAM/ASA J. Uncertain. Quantif. 11 (2023), 1044–1068.

[23] V. Spokoiny, *Mixed Laplace approximation for marginal posterior and Bayesian inference in error-in-operator model*, arXiv:2305.09336v1 (2023).

[24] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Ser. Statist., Springer, New York, 2009.

[25] E. Ullah, T. Mai, A. Rao, R. A. Rossi, and R. Arora, *Machine unlearning via algorithmic stability*, in: Proceedings of Thirty Fourth Conference on Learning Theory, Proc. Mach. Learn. Res. 134 (2021), 4126–4142.

Pavel Ievlev, Timofei Shashkov
Department of Actuarial Science
University of Lausanne
UNIL-Dorigny
1015 Lausanne, Switzerland
E-mail: ievlev.pn@gmail.com
timofei98shashkov@gmail.com

Received 25.3.2026;
accepted 14.7.2026