British Journal of Mathematics & Computer Science
2(4): 198-212, 2012

SCIENCEDOMAIN *international*
www.sciencedomain.org

Kullback–Leibler Divergence of the γ–ordered Normal over t–distribution

Toulias, T-L.[1*] **and Kitsos, C-P.**[1]

[1] *Technological Educational Institute of Athens*
12210 Egaleo, Athens, Greece

Research Article

Received: 24 March 2012
Accepted: 01 June 2012
Published: 08 December 2012

Abstract

The aim of this paper is to evaluate and study the Kullback–Leibler divergence of the γ–ordered Normal distribution, a generalization of Normal distribution emerged from the generalized Fisher's information measure, over the scaled t–distribution. We investigate this evaluation through a series of bounds and approximations while the asymptotic behavior of the divergence is also studied. Moreover, we obtain a generalization of the known Kullback–Leibler information measure between two normal distributions, as well as the K–L divergence between Uniform or Laplace distribution over Normal distribution.

Keywords: Kullback–Liebler divergence; γ–ordered Normal distribution; Scaled t–distribution; Fisher's information measure

2010 Mathematics Subject Classification: 60E15; 62B10; 94A17

1 Introduction

The divergence of information, as a measure of distance between two distributions, attracts special theoretical interest, while various measures have been introduced. The Hellinger distance, being one of them, is an f–divergence measure, see Kamps (1989). Another such measure is the Kullback–Leibler (K–L) divergence (or relative entropy) which is widely used in applied sciences, especially in Signal Processing. It is a significant measure concerning the divergence of the "amount" of information which characterizes certain Input/Output (I/O) systems.

In this paper we study the K–L divergence of a three parameter generalization of the Normal distribution, known as the γ–ordered Normal, over the scaled t–distribution. Emerged from a Logarithm Sobolev Inequality Kitsos and Tavoularis (2009a), the family of γ–ordered Normal distributions provide a "smooth bridging" between Uniform, Normal, Laplace and the degenerated Dirac distributions, see Kitsos and Toulias (2012, 2011). For further reading, see also Kitsos and Toulias

* Tel: +30 210 5385308; fax: +30 210 5385308.
E-mail address: t.toulias@teiath.gr.

(2010a,b). The evaluation of this K–L divergence can be applied in I/O systems that their Input and Output states are described by a wide range of distributions as the γ–ordered Normals $\mathcal{N}_\gamma$ and the scaled t_v–distributions with v degrees of freedom. Both $\mathcal{N}_\gamma$ and t_v families can be considered as two different generalizations of the Normal distribution $\mathcal{N}(\mu, \sigma^2)$ in the sense that $\mathcal{N}_2 = \mathcal{N}$ and $t_\infty = \mathcal{N}$. Therefore, I/O systems with their different states described by distributions "close" to Normal, can be analyzed in terms of their information divergence.

Recall the following definition of the γ–ordered Normal distribution Kitsos and Tavoularis (2009a,b).

Definition 1.1. The random variable X follows the n–variate, γ–order generalized Normal $\mathcal{N}_\gamma^n(\mu, \Sigma)$ with mean vector $\mu \in \mathbb{R}^n$ and positive definite scale matrix $\Sigma \in \mathbb{R}^{n\times n}$, when the density function f_X is of the form

$$f_X(x;\ \mu, \Sigma, \gamma) = C_\gamma^n |\det \Sigma|^{-1/2} \exp\left\{ -\tfrac{\gamma-1}{\gamma} Q(x)^{\frac{\gamma}{2(\gamma-1)}} \right\}, \quad x \in \mathbb{R}^n, \tag{1.1}$$

with Q being the quadratic form $Q(x) = (x-\mu)\,\Sigma^{-1}(x-\mu)^{\mathrm{T}}$. We shall write $X \sim \mathcal{N}_\gamma^n(\mu, \Sigma)$. The normality factor C_γ^n is defined as

$$C_\gamma^n = \pi^{-n/2} \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(n\frac{\gamma-1}{\gamma}+1)} (\tfrac{\gamma-1}{\gamma})^{n\frac{\gamma-1}{\gamma}}. \tag{1.2}$$

Parameter γ is a real number outside the interval $[0, 1]$. Notice that, for $\gamma = 2,\ \mathcal{N}_2^n(\mu, \Sigma)$ coincides with the well known elliptically contoured multivariate Normal distribution.

Recall the K–L divergence of random variables P over Q, defined by Kullback and Leibler (1951) as

$$\mathrm{D}_{KL}(P, Q) = \int_{\mathbb{R}} p(x) \log \frac{p(x)}{q(x)} dx, \tag{1.3}$$

where p and q being the probability densities of P and Q respectively. Moreover, we shall denote by D_{KL} the known K–L information measure between two normally distributed random variables $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z \sim \mathcal{N}(\mu_0, \sigma_0^2)$, which is known to be

$$\mathrm{D}_{KL} = \mathrm{D}_{KL}(X, Z) = \tfrac{1}{2}\left(\log \tfrac{\sigma_0^2}{\sigma^2} - 1 + \tfrac{\sigma^2}{\sigma_0^2}\right) + \tfrac{1}{2\sigma_0^2}|\mu - \mu_0|^2. \tag{1.4}$$

The main result of this paper concerns a series of bounds for the K–L divergence of the univariate γ–ordered Normal distribution over the scaled t_v–distribution. Moreover, there is a particular order of the evaluated bounds, see Theorem 2.1 and Corollary 2.4 in Section 2. For degrees of freedom $v \to \infty$ we come across to a generalization of D_{KL} which is obtained in an exact form, see Theorem 2.2.

Recall that the probability density function f_Y of a t_v–distributed random variable Y with v degrees of freedom, mean $\mu_0 \in \mathbb{R}$, and scale parameter σ_0 (i.e. t_v is the scaled form of usual t_v–distribution), is given by

$$f_Y(x;\ \mu_0, \sigma_0^2, v) = \tfrac{1}{\sigma_0} T_v \left[1 + \tfrac{1}{v}(\tfrac{x-\mu_0}{\sigma_0})^2\right]^{-\frac{v+1}{2}}, \quad \in \mathbb{R}, \tag{1.5}$$

with normality factor

$$T_v = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\,\Gamma(\frac{v}{2})}. \tag{1.6}$$

2 K–L divergence of the γ–ordered Normal over the scaled t–distribution

We investigate the K–L divergence measure $\mathrm{D}_{\gamma,v}$ of the univariate γ–ordered Normal distribution $\mathcal{N}_\gamma(\mu, \sigma^2)$ over the scaled t_v–distribution $t_v(\mu_0, \sigma_0)$. For $X_\gamma \sim \mathcal{N}_\gamma(\mu, \sigma^2)$ and $Y_v \sim t_v(\mu_0, \sigma_0^2)$ where

t_v is the scaled t–distribution we shall denote $\mathrm{D}_{\gamma,v} = \mathrm{D}_{\mathrm{KL}}(X_\gamma, Y_v)$. The following Theorem provides an upper bound for $\mathrm{D}_{\gamma,v}$.

Theorem 2.1. *The K–L divergence $\mathrm{D}_{\gamma,v}$ of the γ–ordered Normal random variable $X_\gamma \sim \mathcal{N}_\gamma(\mu, \sigma^2)$ over the scaled t_v–distributed random variable $Y_v \sim t_v(\mu_0, \sigma_0^2)$, has the following upper bounds, $B_{\gamma,v}$,*

$$\begin{aligned}\mathrm{D}_{\gamma,v} \leq B_{\gamma,v} \quad &= \quad \log C_{\gamma,v} + \log \tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \\ & \tfrac{v+1}{2v}\left[(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}} \frac{\Gamma(3\frac{\gamma-1}{\gamma})}{\Gamma(\frac{\gamma-1}{\gamma})} \tfrac{\sigma^2}{\sigma_0^2} + \left| \tfrac{\mu-\mu_0}{\sigma_0} \right|^2 \right],\end{aligned} \tag{2.1}$$

where

$$C_{\gamma,v} = \frac{\sqrt{v\pi}\,\Gamma(\frac{v}{2})}{2\,\Gamma(\frac{\gamma-1}{\gamma})\,\Gamma(\frac{v+1}{2})} (\tfrac{\gamma}{\gamma-1})^{1/\gamma}. \tag{2.2}$$

Proof. From the definition of the K–L divergence (1.3) and the probability densities f_{X_γ} and f_{Y_v}, as in (1.1) and (1.5) respectively, we obtain

$$\mathrm{D}_{\gamma,v} = \tfrac{1}{\sigma} C_\gamma^1 \left[\left(\log C_{\gamma,v} + \log \tfrac{\sigma_0}{\sigma} \right) I_1 - I_2 + \tfrac{v+1}{2} I_3 \right], \tag{2.3}$$

where $C_{\gamma,v}$ defined as in (2.2) and the integrals I_i, $i = 1, 2, 3$ are given by

$$I_1 = \int_{\mathbb{R}} \exp\left\{ -\tfrac{\gamma-1}{\gamma} \left(\tfrac{1}{\sigma}|x-\mu| \right)^{\frac{\gamma}{\gamma-1}} \right\} dx,$$

$$I_2 = \tfrac{\gamma-1}{\gamma} \int_{\mathbb{R}} (\tfrac{1}{\sigma}|x-\mu|)^{\frac{\gamma}{\gamma-1}} \exp\left\{ -\tfrac{\gamma-1}{\gamma} (\tfrac{1}{\sigma}|x-\mu|)^{\frac{\gamma}{\gamma-1}} \right\} dx, \text{ and}$$

$$I_3 = \int_{\mathbb{R}} \exp\left\{ -\tfrac{\gamma-1}{\gamma} (\tfrac{1}{\sigma}|x-\mu|)^{\frac{\gamma}{\gamma-1}} \right\} \log\left(1 + \tfrac{1}{\sigma_0^2 v} |x-\mu_0|^2 \right) dx.$$

Substituting $z = (\frac{\gamma-1}{\gamma})^{(\gamma-1)/\gamma} \sigma^{-1}(x-\mu)$, we get respectively

$$I_1 = \sigma (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \int_{\mathbb{R}} e^{-|z|^{\frac{\gamma}{\gamma-1}}} dz,$$

$$I_2 = \sigma (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \int_{\mathbb{R}} |z|^{\frac{\gamma}{\gamma-1}} e^{-|z|^{\frac{\gamma}{\gamma-1}}} dz,$$

and

$$I_3 = \sigma (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \int_{\mathbb{R}} e^{-|z|^{\frac{\gamma}{\gamma-1}}} \log\left\{ 1 + \tfrac{1}{v\sigma_0^2} |(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \sigma z + \mu - \mu_0|^2 \right\} dz. \tag{2.4}$$

Recall the known integrals

$$\int_{\mathbb{R}} e^{-|z|^\beta} dz = 2\beta^{-1}\,\Gamma(\tfrac{1}{\beta}) \text{ and } \int_{\mathbb{R}} |z|^\beta e^{-|z|^\beta} dz = \tfrac{1}{\beta} \int_{\mathbb{R}} e^{-|z|^\beta} dz. \tag{2.5}$$

Therefore, the above I_1 and I_2 integrals become

$$I_1 = 2\sigma (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}-1} \Gamma(\tfrac{\gamma-1}{\gamma}) \text{ and } I_2 = \tfrac{\gamma-1}{\gamma} I_1, \tag{2.6}$$

respectively. Thus, (2.3) is reduced to

$$\mathrm{D}_{\gamma,v} = \sigma^{-1} C_\gamma^1 I_1 \left(\log C_{\gamma,v} + \log \tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} \right) + \tfrac{v+1}{2\sigma} C_\gamma^1 I_3.$$

Substituting I_1 from (2.6) and using C_γ^1 from (1.2) then $\mathrm{D}_{\gamma,v}$ can be written as

$$\begin{aligned}\mathrm{D}_{\gamma,v} &= \frac{(\frac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}}}{\Gamma(\frac{\gamma-1}{\gamma}+1)} \cdot \frac{\Gamma(\frac{\gamma-1}{\gamma})}{(\frac{\gamma-1}{\gamma})^{1/\gamma}} \left[\log C_{\gamma,v}^n + p\left(\log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma}\right)\right] + \\ &\quad \frac{v+1}{4\sigma\,\Gamma(\frac{\gamma-1}{\gamma}+1)}(\tfrac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}} I_3,\end{aligned}$$

and applying the gamma function additive identity, we are reduced to

$$\mathrm{D}_{\gamma,v} = \log C_{\gamma,v}^1 + \left(\log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma}\right) + \frac{v+1}{4\sigma\,\Gamma(\frac{\gamma-1}{\gamma})}(\tfrac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}-1} I_3. \tag{2.7}$$

Notice that, the multivariate function in the integral of (2.4) is positive, and so, using the known logarithmic inequality $\log(x+1) \le x,\ x > -1$, relation (2.4) implies

$$I_3 \le (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \tfrac{\sigma}{v\sigma_0^2} \int_{\mathbb{R}} \left|(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}}\sigma z + \mu - \mu_0\right|^2 e^{-|z|^{\frac{\gamma}{\gamma-1}}}\,dz, \tag{2.8}$$

and therefore

$$\begin{aligned} I_3 &\le (\tfrac{\gamma}{\gamma-1})^{3\frac{\gamma-1}{\gamma}} \tfrac{\sigma^3}{v\sigma_0^2} \int_{\mathbb{R}} |z|^2 e^{-|z|^{\frac{\gamma}{\gamma-1}}}\,dz + (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \tfrac{\sigma}{v\sigma_0^2}|\mu-\mu_0|^2 \int_{\mathbb{R}} e^{-|z|^{\frac{\gamma}{\gamma-1}}}\,dz + \\ &\quad 2\tfrac{\sigma^2}{v\sigma_0}(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}}|\mu-\mu_0| \int_{\mathbb{R}} z e^{-|z|^{\frac{\gamma}{\gamma-1}}}\,dz.\end{aligned}$$

The second integral of the above inequality is calculated using the first relation of (2.5) while the third integral is vanished as its integrand is an even function. Thus,

$$\begin{aligned} I_3 &\le \tfrac{2\sigma^3}{v\sigma_0^2}(\tfrac{\gamma}{\gamma-1})^{3\frac{\gamma-1}{\gamma}} \int_{\mathbb{R}_+} z^2 e^{-z^{\frac{\gamma}{\gamma-1}}}\,dz + 2(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}-1} \tfrac{\sigma}{v\sigma_0^2}|\mu-\mu_0|^2\,\Gamma(\tfrac{\gamma-1}{\gamma}) + 0 \\ &= (\tfrac{\gamma}{\gamma-1})^{3\frac{\gamma-1}{\gamma}} \tfrac{2\sigma^3}{3v\sigma_0^2} \int_{\mathbb{R}_+} e^{-z^{\frac{3\gamma}{3(\gamma-1)}}}\,dz^3 + \tfrac{2\sigma}{v\sigma_0^2}|\mu-\mu_0|^2\,\Gamma(\tfrac{\gamma-1}{\gamma})(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}-1}.\end{aligned}$$

Applying the first relation of (2.5), the inequality above is reduced to

$$I_3 \le 2\Gamma(\tfrac{n}{2})(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}-1} \tfrac{\sigma}{v\sigma_0^2}\left[(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}}\sigma^2\,\Gamma(3\tfrac{\gamma-1}{\gamma}) + \Gamma(\tfrac{\gamma-1}{\gamma})|\mu-\mu_0|^2\right].$$

Finally, substituting the above relationship into (2.7), we get

$$\begin{aligned}\mathrm{D}_{\gamma,v} &\le \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \\ &\quad \frac{v+1}{2v\,\Gamma(\frac{\gamma-1}{\gamma})}\left[(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}}\tfrac{\sigma^2}{\sigma_0^2}\,\Gamma(3\tfrac{\gamma-1}{\gamma}) + \sigma_0^{-2}\,\Gamma(\tfrac{\gamma-1}{\gamma})|\mu-\mu_0|^2\right],\end{aligned}$$

and hence (2.1) has been proved. □

We consider now the normal distribution instead of t_v–distribution, i.e. we investigate the limiting case of $v \to \infty$. Then, following Theorem 2.1, we can evaluate the K–L divergence $\mathrm{D}_{\gamma,\infty}$ deriving an exact form for the divergence (without bounds as in Theorem 2.1).

Theorem 2.2. *The K–L divergence $\mathrm{D}_{KL}(X_\gamma, Z) = \mathrm{D}_{\gamma,\infty}$ of the random variable $X_\gamma \sim \mathcal{N}_\gamma(\mu, \sigma^2)$ over the normally distributed random variable $Z \sim \mathcal{N}(\mu_0, \sigma_0^2)$, is given by*

$$\begin{aligned}\mathrm{D}_{\gamma,\infty} &= \log\Big\{ \frac{\sqrt{\pi/2}}{\Gamma(\frac{\gamma-1}{\gamma}+1)} (\tfrac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}} \Big\} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \\ &\quad (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}} \frac{\Gamma(3\frac{\gamma-1}{\gamma})}{2\,\Gamma(\frac{\gamma-1}{\gamma})} (\tfrac{\sigma}{\sigma_0})^2 + \tfrac{1}{2} \left| \tfrac{\mu-\mu_0}{\sigma_0} \right|^2 .\end{aligned} \tag{2.9}$$

Proof. From the proof of Theorem 2.1, substituting (2.4) to (2.7), we get the K–L divergence

$$\mathrm{D}_{\gamma,v} = \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \frac{\frac{\gamma}{\gamma-1} I}{\Gamma(\frac{\gamma-1}{\gamma})}, \tag{2.10}$$

where

$$I = \int_{\mathbb{R}} e^{-|z|^{\frac{\gamma}{\gamma-1}}} \log\Big\{ 1 + \tfrac{1}{v\sigma_0^2} \left| \sigma(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} z + \mu - \mu_0 \right|^2 \Big\}^{v+1} dz.$$

The K–L divergence of $\mathcal{N}_\gamma(\mu, \sigma^2)$ over $\mathcal{N}(\mu_0, \sigma_0^2)$, is the divergence $\mathrm{D}_{\gamma,\infty} = \lim_{v\to\infty} \mathrm{D}_{\gamma,v}$, as the scaled $t_v(\mu_0, \sigma_0^2)$ distribution is, in limit, the normal $\mathcal{N}(\mu_0, \sigma_0^2)$ when $v \to \infty$. The sequence

$$b_v = \frac{\sqrt{v}\,\Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+1}{2}\right)}, \tag{2.11}$$

tends to $\sqrt{2}$ as $v \to \infty$. In particular, $t_\infty(\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2)$ implies that $\lim_{v\to\infty} f_X = f_Z$, where f_X and f_Z are the probability densities of the t_v–distributed random variable $X \sim t_v$ and the normally distributed $Z \sim \mathcal{N}(\mu, \sigma^2)$ respectively. From the definitions (1.5) and (1.1), for $\gamma = 2$, of these densities f_X and f_Z, it is clear that $\lim_{v\to\infty} T_v = C_2^1$, i.e. $\pi^{-1/2} \lim_{v\to\infty} b_v^{-1} = (2\pi)^{-1/2}$, and hence $\lim_{v\to\infty} b_v = \sqrt{2}$. Therefore, substituting the factor $C_{\gamma,v}$ from (2.2) into (2.10), and applying the limit of sequence $b_v \to \sqrt{2}$ together with the fact that $\lim_{v\to\infty}(1+v^{-1})^v = e$, we derive

$$\mathrm{D}_{\gamma,\infty} = \log\Big\{ \frac{\sqrt{\pi}}{\sqrt{2}\,\Gamma(\frac{\gamma-1}{\gamma})} (\tfrac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}-1} \Big\} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \frac{\frac{\gamma}{\gamma-1} I}{\Gamma(\frac{\gamma-1}{\gamma})}, \tag{2.12}$$

where

$$I = \int_{\mathbb{R}} \left| \tfrac{\sigma}{\sigma_0} (\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} z + \tfrac{\mu-\mu_0}{\sigma_0} \right|^2 e^{-|z|^{\frac{\gamma}{\gamma-1}}} dz.$$

Calculating the integral I in (2.12) as the integral in (2.8), we derive

$$I = 2\tfrac{\gamma-1}{\gamma} \left[(\tfrac{\sigma}{\sigma_0})^2 (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}} \Gamma(3\tfrac{\gamma-1}{\gamma}) + \Gamma(\tfrac{\gamma-1}{\gamma}) |\tfrac{\mu-\mu_0}{\sigma_0}|^2 \right],$$

and by substitution in (2.12), we finally obtain (2.9) with the help of the known gamma function additive identity, $\Gamma(x+1) = x\,\Gamma(x),\ x \in \mathbb{R}_+$. □

Notice that, for the "normal" order value $\gamma = 2$, we readily get from (2.9) that $\mathrm{D}_{2,\infty} = \mathrm{D}_{KL}$ as it is expected, with D_{KL} as in (2.3). This is true, as $\mathrm{D}_{\gamma,\infty}$ is reduced to the K–L divergence between two Normal distributions. Therefore, $\mathrm{D}_{\gamma,\infty}$ generalizes the K–L information measure D_{KL} defined in (2.3).

The Uniform and Laplace distributions are members of the family of the γ–ordered Normal distributions, see Kitsos and Toulias (2011, 2012). Therefore, Theorem 2.2 can also provide the K–L divergence of Uniform or Laplace distribution over Normal distribution. Indeed:

Proposition 2.1. *The K–L divergences of the uniformly distributed random variable $U \sim \mathcal{U}(a, b)$ or the Laplace distributed $L \sim \mathcal{L}(\mu, \sigma)$ over the normally distributed $Z \sim \mathcal{N}(\mu_0, \sigma_0^2)$, are given respectively by*

$$\mathrm{D}_{KL}(U, Z) = \mathrm{D}_{1,\infty} = \tfrac{1}{2}\log\tfrac{\pi\sigma_0^2}{b-a} + \tfrac{b-a}{12\sigma_0^2} + \tfrac{1}{8}\sigma_0^{-2}|b + a - 2\mu_0|^2, \tag{2.13}$$

$$\mathrm{D}_{KL}(L, Z) = \mathrm{D}_{\pm\infty,\infty} = \tfrac{1}{2}\log\tfrac{\pi\sigma_0^2}{2\sigma} + \tfrac{\sigma}{\sigma_0^2} - 1 + \sigma_0^{-2}|\mu - \mu_0|^2. \tag{2.14}$$

Proof. Recall that parameter $\gamma \in \mathbb{R} \setminus [0, 1]$. For the limiting order values of $\gamma = 1$ and $\gamma = \pm\infty$ the γ–ordered Normal distribution coincides with the Uniform and Laplace distribution, i.e. we obtain that $\mathcal{N}_1(\mu_{\mathcal{U}}, \sigma_{\mathcal{U}}) = \mathcal{U}(\mu_{\mathcal{U}} - \sigma_{\mathcal{U}}, \mu_{\mathcal{U}} + \sigma_{\mathcal{U}})$ and $\mathcal{N}_{\pm\infty}(\mu, \sigma) = \mathcal{L}(\mu, \sigma)$, Kitsos and Toulias (2011).

(i) For the Laplace case of $\gamma \to \pm\infty$, setting $\frac{\gamma}{\gamma-1} = 1$ into (2.9) we derive (2.14).

(ii) For the uniform case of $\gamma = 1$, it is $\mathrm{D}_{KL}(U, Z) = \mathrm{D}_{1,\infty} = \lim_{\gamma\to 1^+} \mathrm{D}_{\gamma,\infty}$ with $U \in \mathcal{U}(a, b) = \mathrm{N}_1(\mu_{\mathcal{U}}, \sigma_{\mathcal{U}})$. Thus we rewrite (2.9), using the gamma function additive identity $\Gamma(x + 1) = x\,\Gamma(x)$, $x \in \mathbb{R}_+$, in the form

$$\begin{aligned}\mathrm{D}_{\gamma,\infty} \quad = \quad & \log\left\{\frac{\sqrt{\pi/2}}{\Gamma(\frac{\gamma-1}{\gamma}+1)}(\tfrac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}}\right\} + \tfrac{1}{2}(\log\tfrac{\sigma_0^2}{\sigma_{\mathcal{U}}} - \tfrac{\gamma-1}{\gamma}) + \\ & (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}}\frac{\Gamma(3\frac{\gamma-1}{\gamma}+1)}{6\,\Gamma(\frac{\gamma-1}{\gamma}+1)}\tfrac{\sigma_{\mathcal{U}}}{\sigma^2} + \tfrac{1}{2\sigma_{\mathcal{U}}}\|\mu_{\mathcal{U}} - \mu\|^2,\end{aligned}$$

where $\mu_{\mathcal{U}} = \frac{a+b}{2}$ and $\sigma_{\mathcal{U}} = \frac{b-a}{2}$. For $\gamma \to 1^+$ we finally derive (2.13).

Thus, Proposition has been proved □

Corollary 2.3. *When the degrees of freedom $v \in \mathbb{N}$ rise, the bounds $B_{\gamma,v}$ as in* (2.1) *approximate better the K–L divergence $\mathrm{D}_{\gamma,v}$ for all defined $\gamma \in \mathbb{R} \setminus [0, 1]$.*

Proof. Let a_v the sequence $a_v = \frac{v+1}{v}$, $v \in \mathbb{N}$. Then $a_v \to 1$ and $b_v \to \sqrt{2}$ as $v \to \infty$. Considering the bounds $B_{\gamma,v}$ as in (2.1) when $v \to \infty$, it holds that $B_{\gamma,\infty}$ approaches the K–L divergence as in (2.9). Thus, the equality in (2.1), is obtained in limit as $v \to \infty$, i.e. $\mathrm{D}_{\gamma,\infty} = B_{\gamma,\infty}$ and therefore the bounds $B_{\gamma,v}$ approximate better the K–L divergence $\mathrm{D}_{\gamma,v}$ as $v \in \mathbb{N}$ rises, until $B_{\gamma,v}$ coincides with $\mathrm{D}_{\gamma,\infty}$ of Theorem 2.2 for every γ values. □

Figure 1 clarifies the above Corollary 2.3 for $\gamma = 2$.

Corollary 2.4. *The bounds $B_{\gamma,v}$ have a strict descending order converging to $B_{\gamma,\infty} = \mathrm{D}_{\gamma,\infty}$ as v rises, i.e. $B_{\gamma,1} < B_{\gamma,2} < \cdots < B_{\gamma,\infty}$.*

Proof. The sequences $a_v = \frac{v+1}{v}$ and b_v as in (2.11) are descending sequences. As a result, from the form of (2.1), we derive that $B_{\gamma,1} < B_{\gamma,2} < \cdots < B_{\gamma,\infty}$. That is, as t_v–distribution approaches the normal distribution (when $v \to \infty$), the bounds $B_{\gamma,v}$ have a strictly descending order converging to $B_{\gamma,\infty} = \mathrm{D}_{\gamma,\infty}$, see Corollary 2.3. □

In other words, it is shown that when the t_v–distribution approaches the normal distribution, the bounds $B_{2,v}$ of Theorem 2.1 converge, in a descending order, to $\mathrm{D}_{2,\infty}$. Therefore, every $B_{\gamma,v}$ is closer to $\mathrm{D}_{\gamma,\infty}$ than $B_{\gamma,v-1}$. See Figure 1 for an illustration of the above Corollaries 2.4 and 2.3 provided $\gamma = 2$.

Corollary 2.5. *For the normally distributed case, i.e. for $\gamma = 2$, the corresponding bounds $B_{2,v}$ are reduced to*

$$B_{2,v} = \log\frac{\sqrt{\frac{v}{2}}\,\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} + \tfrac{1}{2}\left[\log\tfrac{\sigma_0^2}{\sigma^2} - 1 + \tfrac{v+1}{v}\left(\tfrac{\sigma^2}{\sigma^2} + \sigma_0^{-2}|\mu - \mu_0|^2\right)\right]. \tag{2.15}$$

Moreover, if we let $v \to \infty$, then $\mathrm{D}_{2,\infty} = B_{2,\infty} = \mathrm{D}_{KL}$.

Proof. Considering Theorem 2.1, for the "normal" order value $\gamma = 2$, we readily get (2.15). Moreover, due to limit of (2.11), relation (2.15) implies

$$\lim_{v\to\infty} B_{2,v} = B_{2,\infty} = \tfrac{1}{2}\left(\log\tfrac{\sigma_0^2}{\sigma^2} - 1 + \tfrac{\sigma^2}{\sigma_0^2} + \sigma_0^{-2}|\mu-\mu_0|^2\right),$$

and through Corollary 2.3, $B_{2,\infty} = \mathrm{D}_{KL}$. However, $\mathrm{D}_{2,\infty} = \mathrm{D}_{KL}$, as $\mathrm{D}_{2,\infty}$ being the K–L divergence between two Normals, $\mathcal{N}(\mu_0,\sigma_0^2)$ and $\mathcal{N}(\mu,\sigma^2)$. Therefore, from (2.1), we finally derive $\mathrm{D}_{KL} = \mathrm{D}_{2,\infty} \le B_{2,\infty} = \mathrm{D}_{KL}$. □

Remark 2.1. We investigate now the question of "how good" the bounds $B_{\gamma,v}$ of the K–L divergence $\mathrm{D}_{\gamma,v}$ are. Corollary 2.3 shown that as the degrees of freedom v rises, the better the upper bounds $B_{\gamma,v}$ become approximating the divergence. Moreover, the bounds $B_{\gamma,v}$ also converging to the divergence $\mathrm{D}_{\gamma,v}$ when the scale parameter σ_0 of the t_v–distribution increases. This is due to the use of the logarithm inequality $\log(x+1) \le x$, $x > -1$ utilized in the evaluation of (2.4) (which forms $B_{\gamma,v}$) The fact that the equality in this logarithmic inequality holds for $x = 0$ implies that the logarithm in (2.4) is close to zero as $\sigma_0 \to \infty$. Thus, the inequality in (2.8) become better as σ_0 is getting larger, which leads to better bounds $B_{\gamma,v}$, see also for confirmation Figure 1. Moreover, in case of $\mu = \mu_0$, the bounds $B_{\gamma,v}$ also converge to $\mathrm{D}_{\gamma,v}$ as the scale parameters ratio σ/σ_0 tends to zero. Therefore, the scale parameters behavior is essential for the behavior of the bounds $B_{\gamma,v}$.

This is why the next Theorem investigates the asymptotic behavior of $\mathrm{D}_{\gamma,v}$ with respect the to scale parameters σ and σ_0.

Theorem 2.6. *The K–L divergence of $X_\gamma \sim \mathcal{N}_\gamma(\mu,\sigma^2)$ over a t_v–distributed random variable $Y_v \sim t_v(\mu_0,\sigma_0^2)$ is diverging logarithmically as the shape of Y or X_γ expands or shrinks respectively, i.e. as the value of σ_0 rises or as σ falls. In particular,*

$$\mathrm{D}_{KL}(X_\gamma, Y_v) = \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma}, \tag{2.16}$$

for large values of σ_0, while

$$\mathrm{D}_{KL}(X_\gamma, Y_v) = \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \tfrac{v+1}{2}\log\left\{1 + \tfrac{1}{v\sigma_0^2}|\mu-\mu_0|^2\right\}, \tag{2.17}$$

for quite small values of σ ($\sigma \to 0$).

Proof. It is clear from (2.4) that $I_3 \to 0$ as $\sigma_0 \to \infty$ and therefore, according to (2.7), (2.16) holds for $\sigma_0 \to \infty$, see Figure 3.

Substituting now (2.4) to (2.7) we have that, as $\sigma \to 0$,

$$\mathrm{D}_{\gamma,v} = \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \frac{(v+1)\frac{\gamma}{\gamma-1}}{4\,\Gamma(n\frac{\gamma-1}{\gamma})}\log\left\{1 + \tfrac{1}{v\sigma_0^2}|\mu-\mu_0|^2\right\}\int_{\mathbb{R}^n} e^{-|z|^{\frac{\gamma}{\gamma-1}}}\,dz,$$

and applying the first integral from (2.5) we obtain (2.17), see Figure 3. □

In case of $\mu = \mu_0$, the values of $\mathrm{D}_{KL}(X_\gamma, Y_v)$ diverge logarithmically in the same way, either for large σ_0 or small σ, i.e.

$$\mathrm{D}_{KL}(X_\gamma, Y_v) = \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma}, \text{ as } \tfrac{\sigma}{\sigma_0} \to 0.$$

Also, notice that, asymptotically, $\mathrm{D}_{KL}(X_\gamma, Y_v)$ does not depend on $|\mu-\mu_0|$ for increasing values of σ_0 as in (2.16), i.e. $\mathrm{D}_{KL}(X_\gamma, Y_v)$ values are independent of distance between (the locations) of X_γ and Y_v for large values of σ_0. However, this is not true for the asymptotic behavior of $\mathrm{D}_{KL}(X_\gamma, Y_v)$ when $\sigma \to 0$, as shown in (2.17).

For the "normal" order $\gamma = 2$ the asymptotic behavior of D_{KL} is given in the following Corollary.

Corollary 2.7. *The K–L divergence of a normally distributed $Z \sim \mathcal{N}(\mu, \sigma^2)$ over a t_v–distributed $Y_v \sim t_v(\mu_0, \sigma_0^2)$ is given, asymptotically, by*

$$\mathrm{D}_{2,v} = \begin{cases} \dfrac{2^v \frac{v-2}{2}!}{\sqrt{\pi}(\frac{v+2}{2})^{(v/2)}} + \frac{1}{2}\left(\log \frac{v\sigma_0^2}{2\sigma^2} - 1\right), & v \text{ even}, \\ \log \dfrac{\sqrt{\pi}(\frac{v+1}{2})^{(\frac{v-1}{2})}}{2^{v-1}\frac{v-1}{2}!} + \frac{n}{2}\left(\log \frac{v\sigma_0^2}{2\sigma^2} - 1\right), \quad v > 1, & v \text{ odd}, \\ \log\sqrt{\pi} + \frac{1}{2}\left(\log \frac{\sigma_0^2}{2\sigma^2} - 1\right), & v = 1, \end{cases} \tag{2.18}$$

for large values of σ_0, where $x^{(k)} = x(x+1)...(x+k-1)$, $k \in \mathbb{N} \setminus 0$, $x \in \mathbb{R}$ is the rising factorial (Pochhammer function), while the asymptotic values of $\mathrm{D}_{KL}(Z, Y_v)$ for small enough σ are given by (2.18) *added by the quantity $\frac{v+1}{2}\log\{1 + v^{-1}\sigma_0^{-2}|\mu - \mu_0|^2\}$.*

Proof. Theorem 2.6, for the "normal" order $\gamma = 2$, implies

$$\mathrm{D}_{2,v} = \mathrm{D}_{KL}(Z, Y_v) = \log K_v + \tfrac{1}{2}\left(\log \tfrac{v\sigma_0^2}{2\sigma^2} - 1\right), \tag{2.19}$$

for large σ_0 values, where $K_v = \Gamma(\frac{v}{2})/\Gamma(\frac{v+1}{2})$, $v \in \mathbb{N}$.

(i) Case of $v \in \mathbb{N}$ even. It is $K_v = \frac{v-2}{2}!/\Gamma(\frac{v+1}{2})$ and therefore, applying the known gamma identity

$$\Gamma(k + \tfrac{1}{2}) = \frac{(2k-1)!!}{2^k}\sqrt{\pi} = \frac{(2k)!}{2^{2k}k!}\sqrt{\pi}, \quad k \in \mathbb{N}, \tag{2.20}$$

we get

$$K_v = \frac{2^v \frac{v}{2}! \frac{v-2}{2}!}{\sqrt{\pi} v!},$$

and finally, from the fact that $\frac{(2k)!}{k!} = (k+1)^{(k)}$, $k \in \mathbb{N} \setminus 0$ (implied through the rising factorial notation) we obtain the first branch of (2.18).

(ii) Case of $v \in \mathbb{N}$ odd. From (2.20) and the fact that $\Gamma(\frac{v+1}{2}) = (\frac{v-1}{2})!$, we have

$$K_v = \frac{(v-1)!\sqrt{\pi}}{2^{v-1}(\frac{v-1}{2}!)^2} = \frac{\Gamma(\frac{v-1}{2} + \frac{1}{2})}{\frac{v-1}{2}!} = \frac{\sqrt{\pi}(\frac{v+1}{2})^{(\frac{v-1}{2})}}{2^{v-1}\frac{v-1}{2}!},$$

and hence we obtain, for $v > 1$ and $v = 1$ respectively, the two last branches of (2.18).

Considering (2.17), the asymptotic values of $\mathrm{D}_{KL}(Z, Y_v)$ as $\sigma \to 0$ are given by (2.18) added by $\frac{v+1}{2}\log\{1 + v^{-1}\sigma_0^{-2}|\mu - \mu_0|^2\}$. Figure 3 demonstrate this Corollary. □

A more "compact" form of the upper bound of $\mathrm{D}_{\gamma,v}$, i.e. without the involvement of gamma functions, is given below.

Corollary 2.8. *It holds,*

$$\mathrm{D}_{\gamma,v} \le B_{\gamma,v} < \begin{cases} E_{\gamma,v} + \frac{1}{2}\log\frac{\gamma}{2(\gamma-1)}, & \gamma < 2, \\ E_{\gamma,v}, & \gamma > 2, \end{cases} \tag{2.21}$$

where

$$E_{\gamma,v} = \log\left\{\left(\tfrac{v}{v+1}\right)^{\frac{v-1}{2}}\tfrac{\sigma_0}{\sigma}\right\} + \tfrac{v+1}{2v}\left[\tfrac{1}{\sqrt{3}}\left(\tfrac{3\sqrt{3}}{e}\right)^{2\frac{\gamma-1}{\gamma}}\tfrac{\sigma^2}{\sigma_0^2} + \tfrac{1}{\sigma_0^2}|\mu - \mu_0|^2\right], \tag{2.22}$$

while for $\gamma = 2$,

$$\mathrm{D}_{2,v} \le B_{2,v} < \tfrac{v-1}{2}\log\tfrac{v}{v+1} + \log\tfrac{\sigma_0}{\sigma} + \tfrac{v+1}{2v}\left(\tfrac{\sigma^2}{\sigma_0^2} + \tfrac{1}{\sigma_0^2}|\mu - \mu_0|^2\right). \tag{2.23}$$

Proof. Utilizing the gamma function inequality Chen and Qi (2006),

$$\frac{b^{b-1}}{a^{a-1}}e^{a-b} < \frac{\Gamma(b)}{\Gamma(a)} < \frac{b^{b-\frac{1}{2}}}{a^{a-\frac{1}{2}}}e^{a-b}, \quad 0 < a < b, \tag{2.24}$$

a simpler form of the bound of Theorem 2.1 can be obtained. In particular, applying (2.24) into $\Gamma(3\frac{\gamma-1}{\gamma})/\Gamma(\frac{\gamma-1}{\gamma})$, we get

$$\frac{\Gamma(3\frac{\gamma-1}{\gamma})}{\Gamma(\frac{\gamma-1}{\gamma})} < 3^{\frac{\gamma-1}{\gamma}-\frac{1}{2}}(3\tfrac{\gamma-1}{\gamma})^{2\frac{\gamma-1}{\gamma}}e^{2\frac{1-\gamma}{\gamma}}, \tag{2.25}$$

while for $\Gamma(\frac{v}{2})/\Gamma(\frac{v+1}{2})$, it is

$$\frac{\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} < (\tfrac{v}{v+1})^{\frac{v}{2}-1}\sqrt{\tfrac{2e}{v+1}}. \tag{2.26}$$

We distinguish now the following three cases.

(i) Case $\gamma > 2$. In this case, $\frac{1}{2} < \frac{\gamma-1}{\gamma}$ and therefore, using the inverted ratios of (2.24), we have

$$\frac{\sqrt{\pi}}{\Gamma(\frac{\gamma-1}{\gamma})} = \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{\gamma-1}{\gamma})} < 2e^{\frac{\gamma-2}{2\gamma}}\frac{(\frac{1}{2})^{\frac{1}{2}}}{(\frac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}}}\frac{\gamma-1}{\gamma}. \tag{2.27}$$

(ii) Case $\gamma < 2$. In this case, $\frac{1}{2} > \frac{\gamma-1}{\gamma}$ and therefore using (2.24), we have

$$\frac{\sqrt{\pi}}{\Gamma(\frac{\gamma-1}{\gamma})} = \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{\gamma-1}{\gamma})} < e^{\frac{\gamma-2}{2\gamma}}\frac{(\frac{1}{2})^{\frac{1}{2}}}{(\frac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}}}\sqrt{2\tfrac{\gamma-1}{\gamma}}. \tag{2.28}$$

(iii) Case $\gamma = 2$. Applying (2.26) into (2.15) we get (2.23).

Substituting (2.25), (2.26), (2.27) or (2.25), (2.26), (2.28) in (2.1), we derive, after some calculations, (2.21) and Corollary has been proved. □

Notice that, the new upper bound, as in (2.23), also converges to D_{KL} (when $v \to \infty$), as $B_{2,v}$ does. This is true because (2.23) implies

$$\begin{aligned}\mathrm{D}_{KL} = \lim_{v\to\infty} B_{2,v} &< \lim_{v\to\infty}\{(\tfrac{v+1}{2}-1)\log\tfrac{v}{v+1}\} + \tfrac{1}{2}\left(\log\tfrac{\sigma_0^2}{\sigma^2} + \tfrac{\sigma^2}{\sigma_0^2} + \tfrac{1}{\sigma_0^2}||\mu-\mu_0||^2\right) \\ &= \log\lim_{v\to\infty}(1-\tfrac{1}{1+v})^{\frac{1+v}{2}} + \tfrac{1}{2}\left(\log\tfrac{\sigma_0^2}{\sigma^2} + \tfrac{\sigma^2}{\sigma_0^2} + \tfrac{1}{\sigma_0^2}||\mu-\mu_0||^2\right) = \mathrm{D}_{KL}.\end{aligned}$$

Therefore, the new upper bound, as in (2.23), preserves the same "good" property as $B_{2,v}$ of converging to D_{KL} as $v \to \infty$. Moreover, also preserves the same asymptotic behavior as $B_{2,v}$, of converging to D_{KL}, for $\sigma_0 \to \infty$ or $\sigma \to 0$.

We now state and prove the next Theorem which provides a series of upper and lower bounds for $\mathrm{D}_{\gamma,v}$ through a finite sum expansion of $B_{\gamma,v}$.

Theorem 2.9. *The K–L divergence $\mathrm{D}_{\gamma,v}$ of the γ–ordered Normal distribution $\mathcal{N}_\gamma(\mu,\sigma^2)$ over the scaled $t_v(\mu,\sigma_0^2)$ distribution with the same mean μ, is bounded from*

$$B_{\gamma,v}(2m) \le \mathrm{D}_{\gamma,v} \le B_{\gamma,v}(2m-1), \quad m \in \mathbb{N}\setminus 0, \tag{2.29}$$

where

$$\begin{aligned}B_{\gamma,v}(m) &= \log C_{\gamma,v} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{\gamma-1}{\gamma} + \\ &\frac{v+1}{2\,\Gamma(\frac{\gamma-1}{\gamma})}\sum_{k=0}^{m-1}\tfrac{(-1)^k}{k+1}\left[(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}}\tfrac{\sigma^2}{v\sigma_0^2}\right]^{k+1}\Gamma\left((2k+3)\tfrac{\gamma-1}{\gamma}\right).\end{aligned} \tag{2.30}$$

Proof. From the the known series expansion

$$\log(1+x) = \sum_{k=0}^{\infty} \tfrac{(-1)^k}{k+1} x^{k+1}, \quad |x| < 1, \tag{2.31}$$

it is true that, for the finite sums, we have

$$\sum_{k=0}^{2m} \tfrac{(-1)^k}{k+1} x^{k+1} \leq \log(1+x) \leq \sum_{k=0}^{2m-1} \tfrac{(-1)^k}{k+1} x^{k+1}, \quad x \geq 0,\ m \in \mathbb{N}. \tag{2.32}$$

Thus, from the right–hand inequality of (2.32) the relation (2.4), provided that $\mu = \mu_0$, implies

$$I_3 \leq \sigma(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \left[\sum_{k=0}^{2m-1} \tfrac{(-1)^k}{k+1} \cdot \frac{\sigma^{2(k+1)}}{v^{k+1}\sigma_0^{2(k+1)}} (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}(k+1)} \int_{\mathbb{R}} |z|^{2(k+1)} e^{-|z|^{\frac{\gamma}{\gamma-1}}} dz \right].$$

To calculate the integral of I_3 above, we switch to polar coordinates, i.e.

$$I_3 \leq \sigma(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \left[\sum_{k=0}^{2m-1} \tfrac{(-1)^k}{k+1} \cdot \frac{\sigma^{2(k+1)}}{v^{k+1}\sigma_0^{2(k+1)}} (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}(k+1)} 2 \int_{\mathbb{R}_+} \rho^{2(k+1)} e^{-\rho^{\frac{\gamma}{\gamma-1}}} d\rho \right],$$

and applying the transformation $w = (2k+3)^{-1}\rho^{2k+3}$, we have

$$I_3 \leq 2\sigma(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}} \left[\sum_{k=0}^{2m-1} \frac{(-1)^k (\frac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}(k+1)}}{(k+1)(2k+3)} \cdot \frac{\sigma^{2(k+1)}}{v^{k+1}\sigma_0^{2(k+1)}} \int_{\mathbb{R}_+} e^{-w^{\frac{\gamma}{(\gamma-1)(2k+3)}}} dw \right].$$

Applying the first integral from (2.5), we get,

$$I_3 \leq 2\sigma(\tfrac{\gamma}{\gamma-1})^{\frac{\gamma-1}{\gamma}-1} \sum_{k=0}^{2m-1} \tfrac{(-1)^k}{k+1} \left[(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}} \tfrac{\sigma^2}{v\sigma_0^2} \right]^{k+1} \Gamma\left((2k+3)\tfrac{\gamma-1}{\gamma} \right). \tag{2.33}$$

Finally, substituting I_3 from (2.33) in (2.7), we get the right–hand inequality of (2.29).

Similarly to the above procedure, the left–hand inequality of (2.29) can be proved and therefore (2.29) holds. □

Notice that, the the series of the upper bounds $B_{\gamma,v}(2m-1),\ m \in \mathbb{N}$ generalize the upper bound $B_{\gamma,v}$ as in (2.1) in the sense that $B_{\gamma,v} = B_{\gamma,v}(1)$. Moreover, using the first– and second–termed expression of (2.30) we are reduced to the following Corollary.

Corollary 2.10. *The K–L divergence* $\mathrm{D}_{\gamma,\infty} = \mathrm{D}_{KL}(X_\gamma, Y_v)$*, with* $X_\gamma \sim \mathcal{N}_\gamma(\mu, \sigma^2)$ *and* $Y_v \sim t_v(\mu, \sigma_0^2)$*, can be bounded from*

$$\log M + \tfrac{v+1}{2v\sigma_0^2}\,(VarX_\gamma)k_{\gamma,v} \leq \mathrm{D}_{\gamma,v} \leq \log M + \tfrac{v+1}{2v\sigma_0^2} \operatorname{Var} X_\gamma, \tag{2.34}$$

where

$$k_{\gamma,v} = 1 - \tfrac{1}{2v\sigma_0^2} \operatorname{Var} X_\gamma [\operatorname{Kurt} X_\gamma + 3],$$

and

$$M = \frac{\sqrt{\pi v}\,\Gamma(\frac{v}{2})(\frac{\gamma-1}{e\gamma})^{\frac{\gamma-1}{\gamma}}\sigma_0}{2\,\Gamma(\frac{\gamma-1}{\gamma}+1)\,\Gamma(\frac{v+1}{2})\sigma}.$$

Proof. The variance and kurtosis of the γ–ordered Normal random variable are given respectively by Kitsos and Toulias (2011)

$$\operatorname{Var} X_\gamma = (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}} \frac{\Gamma(3\frac{\gamma-1}{\gamma})}{\Gamma(\frac{\gamma-1}{\gamma})}\sigma^2, \text{ and} \tag{2.35}$$

$$\operatorname{Kurt} X_\gamma = \frac{\Gamma(\frac{\gamma-1}{\gamma})\,\Gamma(5\frac{\gamma-1}{\gamma})}{\Gamma^2(3\frac{\gamma-1}{\gamma})}\sigma^2 - 3. \tag{2.36}$$

For $m = 1$, the bounds as in (2.29) can be expressed, through (2.35) and (2.36), as in (2.34). □

Corollary 2.11. *The bounds $B_{\gamma,v}(m)$ of the K–L divergence $\mathrm{D}_{\gamma,\infty}$ converge to $\mathrm{D}_{\gamma,\infty}$ as the degrees of freedom rises, i.e. $B_{\gamma,\infty}(m) = \mathrm{D}_{\gamma,\infty}$ for every $m \in \mathbb{N} \setminus 0$.*

Proof. Recall b_v from (2.11) with $\lim_{v\to\infty} b_v = \sqrt{2}$, and the descending sequence $c_v = (v+1)/v^{k+1}$, $k \in \mathbb{N}$ for which

$$\lim_{v\to\infty} c_v = \lim_{v\to\infty} \tfrac{v+1}{v^{k+1}} = \begin{cases} 1, & \text{for } k = 0, \\ 0, & \text{for } k \in \mathbb{N} \setminus 0. \end{cases} \tag{2.37}$$

Thus, from (2.30), we derive that $B_{\gamma,\infty}(m) = \lim_{v\to\infty} B_{\gamma,v}(m) = \mathrm{D}_{\gamma,\infty}$, $m \in \mathbb{N} \setminus 0$, see Figure 2. Therefore, the $B_{\gamma,v}(m)$ series of bounds have the same "good quality" as $B_{\gamma,v}$ in Theorem 2.1 regarding the convergence to $\mathrm{D}_{\gamma,v}$ when $v \to \infty$. Therefore, the higher the degrees of freedom v are the better the bounds $B_{\gamma,v}(m)$ become. □

Corollary 2.12. *The K–L divergence $\mathrm{D}_{KL}(Z, Y_v)$ of a normally distributed random variable $Z \sim \mathcal{N}^n(\mu, \sigma^2)$ over the t_v–distributed $Y_v \sim t_v(\mu, \sigma_0^2)$ with the same mean μ, is bounded from*

$$B_{2,v}(2m) \le \mathrm{D}_{KL}(X, Y_v) \le B_{2,v}(2m-1), \quad m \in \mathbb{N} \setminus 0, \tag{2.38}$$

where

$$B_{2,v}(m) = \log \frac{\sqrt{\frac{v}{2}}\,\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} + \log \tfrac{\sigma_0}{\sigma} - \tfrac{1}{2} + \tfrac{v+1}{2} \sum_{k=0}^{m-1} \tfrac{(-1)^k}{k+1} \left(\tfrac{\sigma^2}{2\sigma_0^2}\right)^{k+1} (\tfrac{1}{2})^{(k+1)}, \tag{2.39}$$

while for every $m \in \mathbb{N} \setminus 0$,

$$\lim_{v\to\infty} B_{2,v}(m) = \tfrac{1}{2}\left(\log \tfrac{\sigma^2}{\sigma_0^2} - 1 + \tfrac{\sigma^2}{\sigma_0^2}\right) = \mathrm{D}_{KL}.$$

Proof. For the normal order $\gamma = 2$, (2.30) implies (2.38) where

$$B_{2,v}(m) = \log \frac{\sqrt{\frac{v}{2}}\,\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} + \log \tfrac{\sigma_0}{\sigma} - \tfrac{1}{2} + \tfrac{v+1}{2\sqrt{\pi}} \sum_{k=0}^{m-1} \tfrac{(-1)^k}{k+1} \left(\tfrac{\sigma^2}{2\sigma_0^2}\right)^{k+1} \Gamma(k + \tfrac{3}{2}). \tag{2.40}$$

Utilizing the gamma function additive identity, we have that

$$\Gamma(k + \tfrac{3}{2}) = (k + \tfrac{1}{2})\,\Gamma(k + \tfrac{1}{2}) = (k + \tfrac{1}{2})(k - 1 + \tfrac{1}{2}) \cdots \tfrac{1}{2}\,\Gamma(\tfrac{1}{2}),$$

and through the rising factorial symbol, we get

$$\Gamma(k + \tfrac{3}{2}) = (\tfrac{1}{2})^{(k+1)}\,\Gamma(\tfrac{1}{2}) = (\tfrac{1}{2})^{(k+1)}\sqrt{\pi}, k \in \mathbb{N}.$$

Therefore, from (2.40) we finally derive (2.39).

Considering now $\lim_{v\to\infty} b_v = \sqrt{2}$ with b_v as in (2.11), we get

$$\begin{aligned} \lim_{v\to\infty} B_{2,v}(m) \quad = \quad & \log \tfrac{\sigma_0}{\sigma} - \tfrac{1}{2} + \tfrac{\sigma^2}{2\sigma_0^2}\left(\lim_{v\to\infty} \tfrac{v+1}{v}\right) + \\ & \frac{1}{2\sqrt{\pi}} \left[\sum_{k=1}^{m-1} \tfrac{(-1)^k}{k+1} \left(\tfrac{\sigma^2}{2\sigma_0^2}\right)^{k+1} \Gamma(k + \tfrac{3}{2}) \lim_{v\to\infty} \tfrac{v+1}{v^{k+1}} \right], \end{aligned}$$

and using (2.37), we obtain

$$\lim_{v\to\infty} B_{2,v}(m) = \log\tfrac{\sigma_0}{\sigma} - \tfrac{1}{2} + \tfrac{\sigma^2}{2\sigma_0^2}.$$

This result was expected, as $\mathrm{D}_{2,\infty} = \mathrm{D}_{KL}$ (provided that $\mu = \mu_0$), see also Figure 2. □

Corollary 2.13. *The values of the K–L divergence* $\mathrm{D}_{KL}(Z, Y_v)$ *as in Corollary 2.12 can be approximated as*

$$\mathrm{D}^n_{KL}(X, Y_v) \approx \log\frac{(\frac{\sigma}{v\sigma_0})^{v/2}\,\Gamma(\frac{v}{2})}{2^{v+3/2}\sqrt{\pi}\,\Gamma(\frac{v+1}{2})}. \tag{2.41}$$

Proof. Recall $B_{2,v}(m)$ as in (2.39). Using the fact that $(\frac{1}{2})^{(k+1)} > (\frac{1}{2})^{k+1}$, then

$$\lim_{m\to\infty} B_{2,v}(m) \approx \log\frac{\sqrt{\frac{v}{2}}\,\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} + \log\tfrac{\sigma_0}{\sigma} - \tfrac{1}{2} + \tfrac{v+1}{2}\log\tfrac{\sigma^2}{4v\sigma_0^2},$$

due to (2.31), and from (2.38) we finally derive (2.41). □

The above discussion shows that as the degrees of freedom v rises, the better the upper bounds $B_{\gamma,v}(2m-1)$ or the lower bounds $B_{\gamma,v}(2m)$ become, converging to the K–L divergence $\mathrm{D}_{\gamma,v}$, see Figure 2. Moreover, the bounds $B_{\gamma,v}(m)$, $m \in \mathbb{N}\setminus 0$ are also converging to $\mathrm{D}_{\gamma,v}$ when the scale parameter σ_0 of the t_v–distribution increases, i.e. $\sigma_0 \to \infty$. This is the case because the series expansion in (2.31), as well as the finite sums in (2.32), used for the evaluation of (2.29), satisfy the equality for $x = 0$. This implies that the logarithm in (2.4), used for the proof of (2.29), is close to zero as $\sigma_0 \to \infty$. Thus, the better the inequality in (2.4) becomes, which leads to better bounds $B_{\gamma,v}(m)$, see also Figure 2 for confirmation. In case of $\mu = \mu_0$, similar to the above line of thought, the series of bounds $B_{\gamma,v}(m)$ converge to $\mathrm{D}_{\gamma,v}$ as the scale parameter ratio σ/σ_0 tends to zero.

3 Discussion

This paper studies the K–L divergence $\mathrm{D}_{\gamma,v}$ of the generalization of the Normal distribution, the γ–ordered Normal, over the scaled t_v–distribution, providing a series of bounds which approximate $\mathrm{D}_{\gamma,v}$ under certain conditions. Moreover, a generalization of the K–L information measure D_{KL} as in (2.3) is obtained in Theorem 2.2, which provides also the K–L divergence of a Uniform or Laplace over Normal distribution.

For visualization purposes of the results in Section 2 we present and discuss the following Figures.

1. Figure 1 demonstrates the behavior of the upper bounds $B_{2,v}$ for $v = 1, 2, \ldots, 5$, as evaluated in Corollary 2.4. These $B_{2,v}$ curves are compared with the depicted actual graphs of their corresponding K–L divergences $\mathrm{D}_{KL}(X_2, Y_v)$ for $v = 1, 2, \ldots 5, \infty$, evaluated numerically through (2.3), where $X_2 \sim \mathcal{N}(\mu, 1)$ and $Y_v \sim t_v(\mu, \sigma_0^2)$ are located in same arbitrary mean $\mu \in \mathbb{R}$.

   One can easily notice the strictly descending order of these bounds, i.e. $B_{2,1} < B_{2,2} < \cdots < B_{2,\infty}$ as proved in Corollary 2.3. Also, notice that these bounds are "quite good" approximations of $\mathrm{D}_{2,v}$ and $\mathrm{D}_{2,v}$ for large enough values of the scale parameter σ_0, see Corollary 2.7. Moreover, as the t_v–distribution approaches the normal distribution (i.e. when $v \to \infty$) the corresponding bounds $B_{2,v}$ are getting closer to $\mathrm{D}_{2,v}$ until they all coincide for $v = \infty$, as shown in Corollary 2.4, i.e. $B_{2,\infty} = \mathrm{D}_{2,\infty} = \mathrm{D}_{KL}$.

2. Figure 2 illustrates the behavior of the upper bounds $B_{2,v}(5)$ as well as the lower bounds $B_{2,v}(6)$ for $v = 1, 2, 3$ as evaluated through Corollary 2.12. These bounds are compared to the actual values of $\mathrm{D}_{KL}(X_2, Y_v)$ also depicted for $v = 1, 2, 3, \infty$ (as in Fig. 1), where $X_2 \sim \mathcal{N}(\mu, 1)$ and $Y_v \sim t_v(\mu, \sigma_0^2)$.

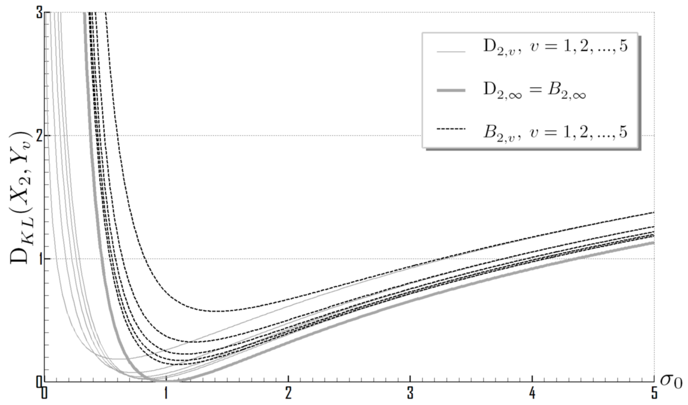

Figure 1: *Graphs of the upper bounds $B_{2,v}$ across σ_0 for various v where $X_2 \sim \mathcal{N}(\mu,1)$ and $Y_v \sim t_v(\mu,\sigma_0^2)$, together with their corresponding K–L divergences $\mathrm{D}_{KL}(X_2,Y_v)$, $\mu \in \mathbb{R}$.*

Notice that these bounds, like $B_{2,v}$ in Fig. 1, are indeed "quite good" approximations of $\mathrm{D}_{2,v}$ for large enough values of the scale parameter σ_0. Moreover, as the t_v–distribution approaches the normal distribution (i.e. $v \to \infty$) the corresponding upper $B_{2,v}(5)$ and lower $B_{2,v}(6)$ bounds are getting closer to $\mathrm{D}_{2,v}$ until they all coincide for $v=\infty$, as proved in Corollary 2.11, i.e. $B_{2,\infty}(m)=\mathrm{D}_{2,\infty}(m)=\mathrm{D}_{KL}$.

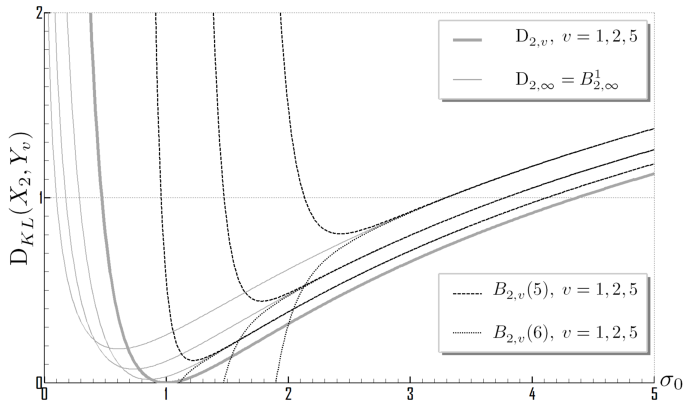

Figure 2: *Graphs of the upper $B_{2,v}(5)$ and lower $B_{2,v}(6)$ bounds across σ_0 for various v, where $X_2 \sim \mathcal{N}(\mu,1)$ and $Y_v \sim t_v(\mu,\sigma_0^2)$, together with their corresponding K–L divergences $\mathrm{D}_{KL}(X_2,Y_v)$, $\mu \in \mathbb{R}$.*

3. Figure 3 illustrates the asymptotic behavior of $\mathrm{D}_{2,v}(X_2,Y_v)$ for large σ_0 (left–side) or small σ (right–side) with various degrees of freedom v through the depiction of $A_{2,v}(X_2,Y_v)$ as in (**??**), together with their corresponding actual K–L divergences $\mathrm{D}_{2,v}(X_2,Y_v)$ for any $\mu \in \mathbb{R}$. The random variables $X_2 \sim \mathcal{N}(\mu,1)$ and $Y_v \sim t_v(\mu,\sigma_0^2)$ were used in Fig. 3a while $X_2 \sim \mathcal{N}(\mu,\sigma^2)$ and $Y_v \sim t_v(\mu,1)$ (Y_v is the usual, not scaled, t_v–distribution) were used in Fig. 3b.
4. Figure 4 illustrates a series of the $B_{2,v=1,5}(m)$ upper bounds in various m–termed forms as in (2.30). For the $v=1$ case see Fig. 4a while for the $v=5$ case see Fig. 4b.

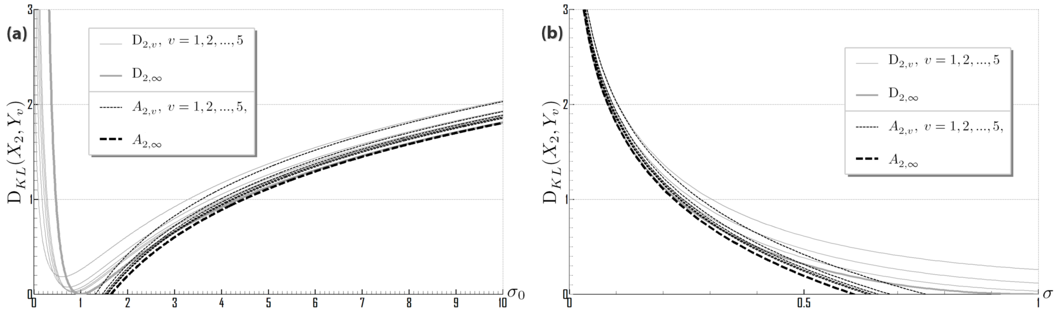

Figure 3: *Graphs of $A_{2,1}(X_2, Y_v)$ across σ_0 (Fig. 3a) and σ (Fig. 3b), together with their corresponding K–L divergences $D_{2,1}(X_2, Y_v)$, $\mu \in \mathbb{R}$.*

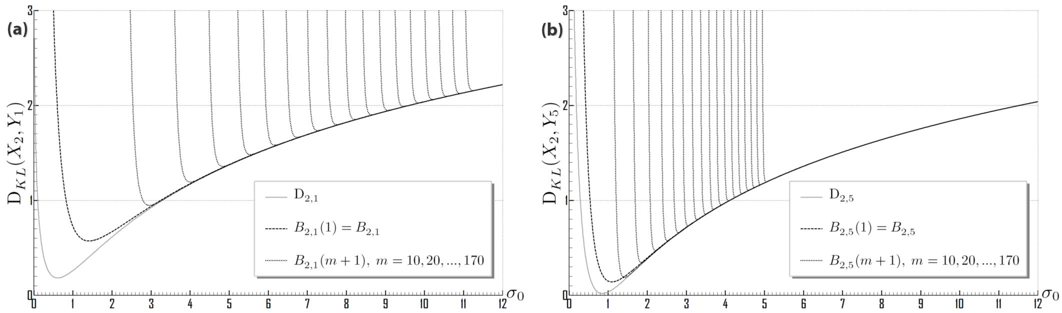

Figure 4: *Graphs of the upper bounds $B_{2,v}(m)$ across σ_0 for $v = 1$ (Fig. 4a) and $v = 5$ (Fig. 4b), evaluated with various odd m–terms, where $X_2 \sim \mathcal{N}(\mu, 1)$ and $Y_v \sim t_v(\mu, \sigma_0^2)$, together with their corresponding K–L divergences $D_{2,1}(X_2, Y_v)$, $\mu \in \mathbb{R}$.*

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed equally and significantly in this research work. All authors read and approved the final manuscript.

Acknowledgment

The authors would like to thank the referees for their useful comments which helped us to provide the final version of this paper.

References

Chen, C-P., Qi, F. (2006). Logarithmically complete monotonic functions relating to the gamma function. J. Math. Appl. 321, 405–411.

Kamps, U. (1989). Hellinger distances and α–entropy in a one–parameter class of density functions. Statistical Papers 30, 263–269.

Kitsos, C-P., Tavoularis, N-K. (2009) Logarithmic Sobolev inequalities for information measures. IEEE Trans. Inform. Theory 55(6), 2554–2561.

Kitsos, C-P., Tavoularis, N-K. (2009). New entropy type information measures. In: Proceedings of the conference ITI2009-information technology interfaces-Cavtat, Croatia: June 22–25.

Kitsos, C-P., Toulias, T-L. (2012). On the multivariate γ–ordered normal distribution. To appear in Far East J. Theor. Stat.

Kitsos, C-P., Toulias, T-L. (2011). On the family of the γ–ordered normal distributions. Far East J. Theor. Stat. 35(2), 95–114.

Kitsos, C-P., Toulias, T-L. (2010). New information measures for the generalized normal distribution. Information 1, 13–27.

Kitsos, C-P., Toulias, T-L. (2010). Evaluating information measures for the γ-order Multivariate Gaussian. In: Proceedings of the conference PCI2010-Panhellenic Conference on Informatics-University of Peloponnese, Tripoli, Greece: September 10–12.

Kullback, S., Leibler, A. (1951). On information and sufficiency. Ann. Math. Stat. 22, 79–86.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy and paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=165&id=6&aid=768