KULLBACK-LEIBLER DISTANCE BETWEEN COMPLEX GENERALIZED GAUSSIAN DISTRIBUTIONS

Corina Nafornita[1], Yannick Berthoumieu[2], Ioan Nafornita[1], Alexandru Isar[1]

[1]Politehnica University of Timisoara, Romania, corina.nafornita@gmail.com
[2] IMS- Dpt LAPS - UMR 5218 CNRS, ENSEIRB - Université de Bordeaux, France

ABSTRACT

In texture classification, feature extraction can be made in a transform domain. A possibility to preserve the translation invariance is to use a complex transform like the Hyperanalytic Wavelet transform. It exhibits a circularly symmetric density function for subband coefficients so it can be modeled by a particular form of the complex generalized Gaussian (CGGD) distribution function. The Kullback-Leibler (KL) divergence, or distance, can be used to measure the similarity between subbands density function. We derive in this paper a closed-form expression for the KL divergence between two complex generalized Gaussian distributions.

Index Terms**—** Kullback-Leibler distance, divergence, Complex Generalized Gaussian Distribution

1. INTRODUCTION

In probability and information theory, the Kullback–Leibler (KL) divergence is a non-symmetric measure of the difference between two probability density functions (pdf), p and q. This is defined as [1]:

$$D_{KL}\left(p\|q\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p\left(x,y\right)\log\frac{p\left(x,y\right)}{q\left(x,y\right)}dxdy \tag{1}$$

If the two pdfs are the same (p=q), the divergence is null. The KL distance is used as a similarity measure between textures, which makes it useful for texture classification [2]. In [2], the authors deal with computation of KL divergence for statistics of real wavelet subband coefficients. A wavelet subband is modeled using the generalized Gaussian distribution (GGD). Based on this model, hyperparameters of the coefficients pdf from each subband are estimated. The KL divergence is computed between the pdf of subbands for two compared textures.

If this classification is made using a complex wavelet transform, we need a complex model and the closed-form for the KL divergence.

The generalization for the GGD model in the complex case was proposed by Novey and Adali which approximates the pdf based on a histogram [3]. The computation problem for the distance between two pdf for complex variables was also discussed by Verdoolaege [4]. He established equations for geodesics in probability space. Unfortunately, these relations are not usable at this moment.

Because the hyperanalytic wavelet transform (HWT) produces complex coefficients with a circular distribution we have studied the simpler problem of KL divergence for such distributions [5]. We derive in this paper a closed-form for the KL divergence of pairs of CGGD random variables and we study its sensitivity with the shape parameter.

The paper has the following structure. In section 2 we give the definition of HWT and its main statistical properties. Section 3 briefly presents the CGGD [3] and we explain why we chose this model for HWT. Section 4 presents the closed-form of the KL divergence of two CGGD. The sensitivity of this KL divergence with the parameters of the CGGDs is analyzed as well. Conclusions are presented in the last section.

2. HWT TRANSFORM

In [5] a new complex wavelet transform was proposed, based on the hypercomplex mother wavelet $\psi_a\left(x,y\right)$ associated to a real mother wavelet $\psi\left(x,y\right)$:

$$\begin{aligned}\psi_a\left(x,y\right)=\psi\left(x,y\right)+i\mathcal{H}_x\left\{\psi\left(x,y\right)\right\}+\\+j\mathcal{H}_y\left\{\psi\left(x,y\right)\right\}+k\mathcal{H}_x\left\{\mathcal{H}_y\left\{\psi\left(x,y\right)\right\}\right\}\end{aligned} \tag{2}$$

where $i^2=j^2=-k^2=-1$, and $ij=ji=k$ [6], $\mathcal{H}_x$ is the Hilbert transform computed across rows and $\mathcal{H}_y$ across columns. The HWT of the image $f\left(x,y\right)$ is:

$$HWT\left\{f\left(x,y\right)\right\}=\left\langle f\left(x,y\right),\psi_a\left(x,y\right)\right\rangle. \tag{3}$$

This is computed using the 2D discrete wavelet transform (2D-DWT) of its associated hypercomplex image, f_a:

$$HWT\left\{f\left(x,y\right)\right\}=DWT\left\{f_a\left(x,y\right)\right\} \tag{4}$$

where f_a is defined as

$$f_a\left(x,y\right)=f\left(x,y\right)+i\mathcal{H}_x\left\{f\left(x,y\right)\right\}+$$

$$+j\mathcal{H}_y\{f(x,y)\}+k\mathcal{H}_x\{\mathcal{H}_y\{f(x,y)\}\}$$

This means that HWT uses four trees, implemented by 2D-DWT, being adequate to a multi-wavelet environment [5]. The HWT identifies six orientations, 3 positive and 3 negative, ±atan(1/2), ±π/4 and ±atan(2):

$$z_{\pm} = z_{\pm R} + jz_{\pm I} \tag{5}$$

A problem of interest is the statistical modeling of the HWT coefficients. For input random processes, random variables as Z, can be associated to the HWT coefficients z.

The coefficients have zero mean, the cross-correlation between their real and imaginary parts is zero and the variances of their real and imaginary parts are estimated to be the same, $\sigma_R^2 = \sigma_I^2 = \sigma^2/2$, for any second order stationary bivariate input random process [7]. Therefore, we considered the repartitions of the random variables $Z_{\pm}$ corresponding to the HWT coefficients $z_{\pm}$ to be like circularly symmetric. The cross-correlation matrix is:

$$\mathbf{C}_b = E\{\mathbf{Z}_b\mathbf{Z}_b^T\} = \begin{bmatrix} \sigma^2/2 & 0 \\ 0 & \sigma^2/2 \end{bmatrix} \tag{6}$$

where $\mathbf{Z}_b=[Z_R, Z_I]^T$ is the bivariate vector of the real and imaginary parts of the HWT coefficients. The augmented form: $\mathbf{Z}_a=[Z, Z^*]^T$ [3] can also be used.

3. CGGD

For a complex generalized Gaussian distribution, CGGD, where the bivariate random vector is $\mathbf{Z}_b$ and the augmented vector is $\mathbf{Z}_a$ [3], the general form of the bivariate covariance matrix is:

$$\mathbf{C}_b = E\{\mathbf{Z}_b\mathbf{Z}_b^T\} = \begin{bmatrix} \sigma_R^2 & \rho \\ \rho & \sigma_I^2 \end{bmatrix} \tag{7}$$

where $\rho = E\{Z_R Z_I\}$ is the cross-correlation between the real and imaginary part. The augmented covariance matrix is established by Novey and Adali as:

$$\mathbf{C}_a = E\{\mathbf{Z}_a\mathbf{Z}_a^H\} = \begin{bmatrix} \sigma_R^2+\sigma_I^2 & (\sigma_R^2-\sigma_I^2)+2j\rho \\ (\sigma_R^2-\sigma_I^2)-2j\rho & \sigma_R^2+\sigma_I^2 \end{bmatrix} \tag{8}$$

The probability density function generalizes the GGD family of densities,

$$p_{\mathrm{X}}(x;\sigma,c) = \frac{c}{2\sigma\Gamma\left(\frac{1}{c}\right)}\exp\left\{-\left(\frac{|x|}{\sigma}\right)^c\right\} \tag{9}$$

where $\Gamma(\cdot)$ is the gamma function, σ is the scale parameter, and c is the shape parameter. The generalized probability density function for the augmented vector is [3]:

$$p_{\mathrm{V}a}(\mathbf{v}_a) = \frac{\beta(c)}{\sqrt{|\mathbf{C}_a|}}\exp\left\{-\left[\eta(c)\left(\mathbf{v}_a^H\mathbf{C}_a^{-1}\mathbf{v}_a\right)\right]^c\right\} \tag{10}$$

where $\mathbf{v}_a = \frac{\mathbf{z}_a}{\sqrt{\Gamma(2/c)/\Gamma(1/c)}}$, $\beta(c) = \frac{c\Gamma(2/c)}{\pi\Gamma^2(1/c)}$ and $\eta(c) = \frac{\Gamma(2/c)}{2\Gamma(1/c)}$. In [3] a Matlab program is presented which gives the ML estimation for the vector $\phi = \left[\sigma_R^2, \sigma_I^2, \rho, c\right]^T$. This means we can have the ML estimation for the shape parameter c and the matrices $\mathbf{C}_b$ and $\mathbf{C}_a$. We show in the following the importance of the quality of this estimation.

4. KULLBACK-LEIBLER DIVERGENCE FOR CGGD

In the case of circular vectors, with $\sigma_R^2 = \sigma_I^2 = \sigma^2/2$ and $\rho = 0$, which corresponds to the HWT coefficients of any bivariate stationary random process [7], starting from the augmented pdf in (10), the bivariate pdf is:

$$p(x,y) = \frac{\beta(c)}{\sigma^2}\exp\left\{-\left(\frac{\Gamma(2/c)}{\Gamma(1/c)}\right)^c\left(\frac{x^2+y^2}{\sigma^2}\right)^c\right\} \tag{11}$$

For the pdf having the shape parameters c_1, c_2 and the variances $\sigma_1^2 = \sigma_2^2$ using relationship (11) and the definition in (1) we obtain the Kullback-Leibler distance:

$$D_{KL}(p_1\|p_2) = \ln\left(\frac{c_1}{c_2}\frac{\sigma_2^2}{\sigma_1^2}\frac{\Gamma(2/c_1)\Gamma^2(1/c_2)}{\Gamma(2/c_2)\Gamma^2(1/c_1)}\right) - \frac{1}{c_1} + \frac{1}{\Gamma(1/c_1)}\left(\frac{\sigma_1^2}{\sigma_2^2}\frac{\Gamma(2/c_1)\Gamma(1/c_2)}{\Gamma(1/c_1)\Gamma(2/c_2)}\right)^{c_2}\Gamma\left(\frac{1}{c_1}+\frac{c_2}{c_1}\right) \tag{12}$$

The proof of this relation can be found in Appendix.

We plot the KL distance between p_1 and p_2, for $\sigma_1 = \sigma_2$. In Fig.1, the shape parameter for p_2, that is c_2, is fixed, with values 0.3, 0.5, 1, 1.5 and 2. The shape parameter for p_1, that is c_1, varies from 0.2 to 2. In Fig.2, the shape parameter for p_1, c_1 is fixed, with values 0.3, 0.5, 1, 1.5 and 2. The shape parameter for p_2, that is c_2, varies from 0.2 to 2.

It is essential for any classification that the distance between the two pdf to be as discriminant as possible. In other words, if c_1 and c_2 are very close then KL should be close to zero, and if they have different values, this distance should be as high as possible.

It can be observed, analyzing Fig. 1 and Fig. 2 that the KL becomes zero if c_1=c_2 and σ_1=σ_2. These parameters are not a priori known in textures classification applications and they must be estimated. The success of the classification depends on the quality of the estimators used. For an efficient classification, it is necessary that the speed of variation of the curves in Fig. 1 and Fig. 2 around their intersections with the line expressed by the equation D_{KL}=0, to be as high as possible.

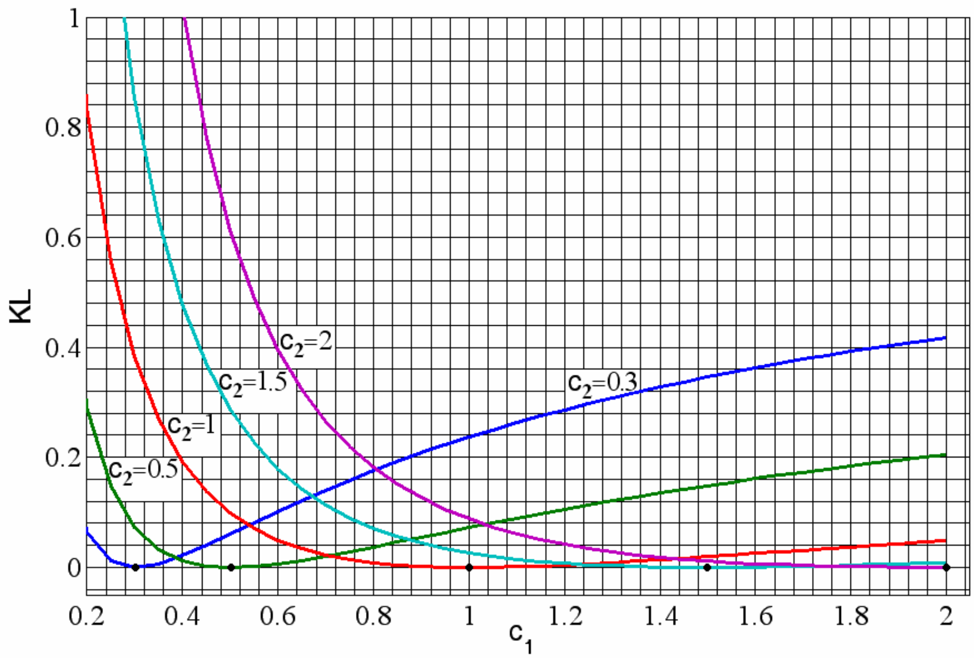

Fig. 1. KL distance between p_1 and p_2 ($\sigma_1 = \sigma_2$). The shape parameter for p_2, c_2 is fixed, with values 0.3, 0.5, 1, 1.5 and 2. The shape parameter for p_1, that is c_1, varies from 0.2 to 2.

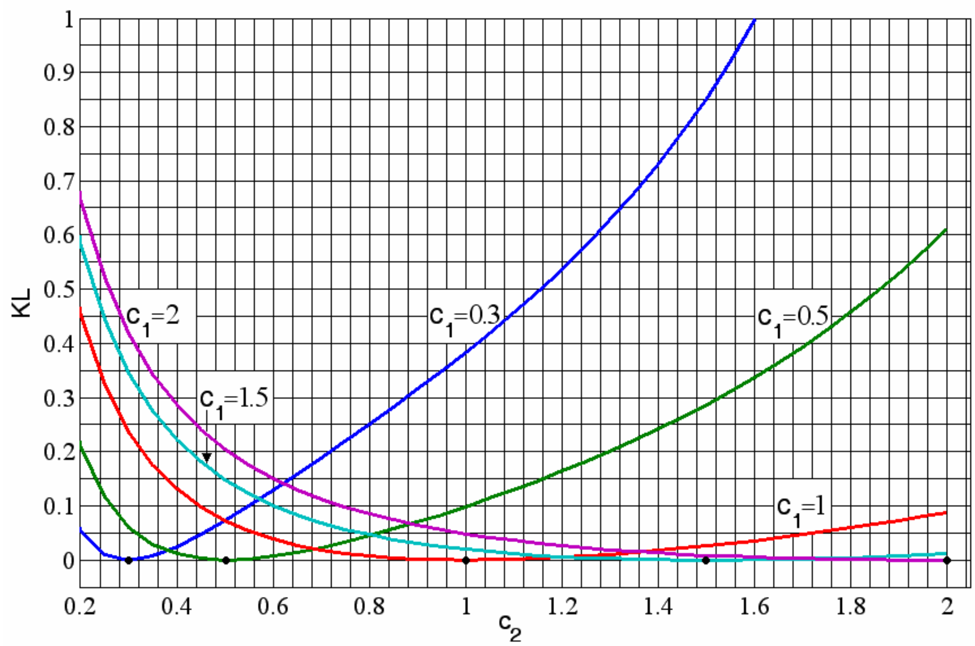

Fig. 2. KL distance between p_1 and p_2 ($\sigma_1 = \sigma_2$). The shape parameter for p_1, c_1 is fixed, with values 0.3, 0.5, 1, 1.5 and 2. The shape parameter for p_2, that is c_2, varies from 0.2 to 2.

For the VisTex database [8], using 40 images subdivided in 16 subimages each, resulting in 640 smaller images, we have repeated the estimation of the shape parameter c and of the covariance matrix $\mathbf{C}_a$, using the programs presented in [3]. This was done in the HWT domain, using one decomposition level and Daubechies-3 mother wavelet. We have noticed that the shape parameter varies in the range 0.1÷5 but its values around 0.5 appear more frequently.

From Fig.1 it is easily noticeable that the KL distance varies only slightly for values of c_1 between 0.8 and 1.2. It is interesting that it responds better around the value c_1=0.25. The KL distance is more sensitive for the plot c_2=0.3 than for Gaussian case (c_2=1).

For Fig.2, where we plotted KL distance with c_1 fixed, the best case is for c_1=0.3, as opposed to the case of c_1=1 (Gaussian case). The KL distance varies only slightly for example in the range c_2 of 0.5÷1.5. As expected, the KL distance is non-symmetric with respect to c_1 and c_2.

5. CONCLUSIONS

In texture classification, when using a complex transform such as the HWT, modeled by the CGGD distribution, the KL distance can be used to measure the similarity between subband density functions. This is not always satisfactory because there are intervals where KL distance varies only slightly despite the fact that the two pdfs are very different. It would be useful in the future to study more measures for texture classification.

6. ACKNOWLEDGMENTS

This paper was developed in the framework of bilateral program Brancusi 510/31.03.2011 “Classification de textures fondée sur la théorie des ondelettes hyper-analytiques et les copules”.

7. REFERENCES

[1] S. Kullback and R.A. Leibler, “On Information and Sufficiency”, *Annals of Mathematical Statistics*, 22 (1), pp. 79–86, 1951.

[2] M. N. Do and M. Vetterli, “Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance”, *IEEE Trans. Image Processing*, vol. 11, pp. 146-158, Feb. 2002.

[3] M. Novey, T. Adali, and A. Roy, “A complex generalized Gaussian distribution-Characterization, generation, and estimation”, *IEEE Trans. Signal Processing*, vol. 58, no. 3, part. 1, pp. 1427-1433, March 2010.

[4] G. Verdoolaege, S. De Backer, P. Scheunders, “Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models”, *ICIP'2008*, pp.169-172.

[5] I. Firoiu, C. Nafornita, J.-M. Boucher, A. Isar, “Image Denoising Using a New Implementation of the Hyperanalytic Wavelet Transform”, *IEEE Trans. on Instrumentation and Measurement*, Aug. 2009, vol. 58, no. 8, pp. 2410-2416.

[6] C. Davenport, Commutative Hypercomplex Mathematics, http://home.comcast.net/~cmdaven/hyprcplx.htm

[7] C. Nafornita, I. Firoiu, D. Isar, J.-M. Boucher, A. Isar, A Second Order Statistical Analysis of the Hyperanalytic Wavelet Transform, *Proc. 9th Int. Symp. on Electronics and Telecommunications*, ISETC 2010, Timisoara, Romania, Nov. 2010, pp. 311-314.

[8] MIT Vision and modeling group, Vision texture, Available online: http://vismod.media.mit.edu/vismod/

APPENDIX

We compute the KL distance for the CGGD model, in the circular case. The probability density function is:

$$\begin{aligned} p(x,y) &= \frac{\beta(c)}{\sigma^2}\exp\left\{-\left(\frac{\Gamma(2/c)}{\Gamma(1/c)}\right)^c\left(\frac{x^2+y^2}{\sigma^2}\right)^c\right\} \\ &= A\exp\left\{-\left(\frac{x^2+y^2}{B^2}\right)^c\right\} \end{aligned} \tag{A.1}$$

where x and y are the real and imaginary components, and

$$A=\frac{c\Gamma\left(\frac{2}{c}\right)}{\pi\Gamma^2\left(\frac{1}{c}\right)\sigma^2} \quad \text{and} \quad B^2=\frac{\sigma^2\Gamma\left(\frac{1}{c}\right)}{\Gamma\left(\frac{2}{c}\right)} \tag{A.2}$$

We compare two pdf:

$$p_1 = A_1\exp\left\{-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}\right\} \tag{A.3}$$

and

$$p_2 = A_2\exp\left\{-\left(\frac{x^2+y^2}{B_2^2}\right)^{c_2}\right\} \tag{A.4}$$

We start from the KL distance definition:

$$D_{KL}\left(p_1\|p_2\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p_1(x,y)\log\frac{p_1(x,y)}{p_2(x,y)}dxdy \tag{A.5}$$

First we have:

$$\ln\frac{p_1}{p_2}=\ln\frac{A_1}{A_2}-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}+\left(\frac{x^2+y^2}{B_2^2}\right)^{c_2} \tag{A.6}$$

The integrand is then:

$$\begin{aligned} p_1\ln\frac{p_1}{p_2} &= A_1\exp\left\{-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}\right\}\cdot \\ &\cdot\left\{\ln\frac{A_1}{A_2}-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}+\left(\frac{x^2+y^2}{B_2^2}\right)^{c_2}\right\} \end{aligned} \tag{A.7}$$

The KL distance can be written as a sum of three terms, I_1, I_2 and I_3:

$$D_{KL}\left(p_1\|p_2\right)=I_1+I_2-I_3 \tag{A.8}$$

The first term is:

$$I_1 = A_1\int_{-\infty}^{\infty}\int\exp\left\{-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}\right\}\ln\frac{A_1}{A_2}dxdy$$

$$I_1 = A_1 \int_0^{2\pi}\int_0^{\infty} \exp\left\{-\left(\frac{r^2}{B_1^2}\right)^{c_1}\right\} \ln\frac{A_1}{A_2}\, r dr d\varphi$$

$$= 2\pi A_1 \ln\frac{A_1}{A_2}\int_0^{\infty} \exp\left\{-\left(\frac{r^2}{B_1^2}\right)^{c_1}\right\} r dr \qquad \text{(A.9)}$$

Because:

$$\frac{r^2}{B_1^2} = t^{\frac{1}{c_1}} \quad \Rightarrow \quad 2rdr = \frac{B_1^2}{c_1} t^{\frac{1}{c_1}-1} dt \qquad \text{(A.10)}$$

we obtain:

$$I_1 = \pi A_1 \ln\frac{A_1}{A_2}\frac{B_1^2}{c_1}\int_0^{\infty} t^{\frac{1}{c_1}-1} e^{-t} dt$$

$$= \pi A_1 \left(\ln\frac{A_1}{A_2}\right)\frac{B_1^2}{c_1}\Gamma\left(\frac{1}{c_1}\right) \qquad \text{(A.11)}$$

In the same manner, we have:

$$I_2 = -A_1 \iint_{\mathbb{R}} \left(\frac{x^2+y^2}{B_1^2}\right)^{c_1} \exp\left\{-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}\right\} dxdy$$

$$= -2\pi A_1 \int_0^{\infty}\left(\frac{r^2}{B_1^2}\right)^{c_1} \exp\left\{-\left(\frac{r^2}{B_1^2}\right)^{c_1}\right\} rdr \qquad \text{(A.12)}$$

$$= -\pi A_1 \frac{B_1^2}{c_1}\Gamma\left(1+\frac{1}{c_1}\right)$$

and

$$I_3 = A_1 \iint_{\mathbb{R}} \left(\frac{x^2+y^2}{B_2^2}\right)^{c_2} \exp\left\{-\left(\frac{x^2+y^2}{B_1^2}\right)^{c_1}\right\} dxdy$$

$$= 2\pi A_1 \int_0^{\infty}\left(\frac{r^2}{B_2^2}\right)^{c_2} \exp\left\{-\left(\frac{r^2}{B_1^2}\right)^{c_1}\right\} rdr$$

$$I_3 = \pi A_1 \int_0^{\infty} t^{\frac{c_2}{c_1}} \left(\frac{B_1^2}{B_2^2}\right)^{c_2} e^{-t} \frac{B_1^2}{c_1} t^{\frac{1}{c_1}-1} dt \qquad \text{(A.13)}$$

$$I_3 = \pi A_1 \left(\frac{B_1^2}{B_2^2}\right)^{c_2} \frac{B_1^2}{c_1}\int_0^{\infty} t^{\frac{c_2}{c_1}+\frac{1}{c_1}-1} e^{-t} tdt$$

$$= \pi A_1 \left(\frac{B_1^2}{B_2^2}\right)^{c_2} \frac{B_1^2}{c_1}\Gamma\left(\frac{1+c_2}{c_1}\right)$$

The distance becomes:

$$D_{KL}\left(p_1 \| p_2\right) = \pi A_1 \left(\ln\frac{A_1}{A_2}\right)\frac{B_1^2}{c_1}\Gamma\left(\frac{1}{c_1}\right)$$

$$-\pi A_1 \frac{B_1^2}{c_1}\Gamma\left(1+\frac{1}{c_1}\right) \qquad \text{(A.14)}$$

$$+\pi A_1 \left(\frac{B_1^2}{B_2^2}\right)^{c_2}\frac{B_1^2}{c_1}\Gamma\left(\frac{1+c_2}{c_1}\right)$$

where:

$$A_i = \frac{c_i \Gamma\left(\frac{2}{c_i}\right)}{\pi \Gamma^2\left(\frac{1}{c_i}\right)\sigma_i^2}; \quad B_i^2 = \frac{\sigma_i^2 \Gamma\left(\frac{1}{c_i}\right)}{\Gamma\left(\frac{2}{c_i}\right)} \quad i = 1, 2 \qquad \text{(A.15)}$$

It results that:

$$D_{KL}\left(p_1 \| p_2\right) = \ln\left(\frac{c_1}{c_2}\frac{\sigma_2^2}{\sigma_1^2}\frac{\Gamma(2/c_1)\Gamma^2(1/c_2)}{\Gamma(2/c_2)\Gamma^2(1/c_1)}\right)$$

$$-\frac{1}{c_1}$$

$$+\frac{1}{\Gamma(1/c_1)}\left(\frac{\sigma_1^2}{\sigma_2^2}\frac{\Gamma(2/c_1)\Gamma(1/c_2)}{\Gamma(1/c_1)\Gamma(2/c_2)}\right)^{c_2}\Gamma\left(\frac{1}{c_1}+\frac{c_2}{c_1}\right)$$

(A.16)

We took into account that:

$$\Gamma\left(1+\frac{1}{c_1}\right) = \frac{1}{c_1}\Gamma\left(\frac{1}{c_1}\right) \qquad \text{(A.17)}$$

We verify that the distance is correct, for

$$c = c_1 = c_2; \qquad \sigma = \sigma_1 = \sigma_2$$

it should be zero:

$$D_{KL}\left(p \| p\right) = \ln\left(\frac{c}{c}\frac{\sigma^2}{\sigma^2}\frac{\Gamma(2/c)\Gamma^2(1/c)}{\Gamma(2/c)\Gamma^2(1/c)}\right) - \frac{1}{c}$$

$$+\frac{1}{\Gamma(1/c)}\left(\frac{\sigma^2}{\sigma^2}\frac{\Gamma(2/c)\Gamma(1/c)}{\Gamma(1/c)\Gamma(2/c)}\right)^{c}\Gamma\left(\frac{1}{c}+\frac{c}{c}\right) \qquad \text{(A.18)}$$

$$= -\frac{1}{c} + \frac{1}{\Gamma(1/c)}\frac{1}{c}\Gamma\left(\frac{1}{c}\right) = 0$$