Divisive Normalization: Justification and Effectiveness as Efficient Coding Transform

Supplementary Materials

1 Introduction

This document includes formal proofs of in NIPS 2010 submission titled *"Divisive Normalization: Justification and Effectiveness as Efficient Coding Transform"*. For convenience, we keep the same order of lemmas as in the main submission, and mark supporting results as additional facts.

2 Basic Notations

$\mathbf{x} \in R^d$	d-dimensional vector data
$x_i \in R$	ith element of vector $\mathbf{x}$
$\mathbf{x}_{\setminus i} \in R^{d-1}$	vector formed by excluding the i element from $\mathbf{x}$
$(\cdot)_+$	Rectifying function, $(a)_+ = a$ for $a \geq 0$ and $(a)_+ = 0$ for $a < 0$
$\Gamma(\beta)$	Gamma function, $\Gamma(\beta) = \int_0^\infty u^{\beta-1} \exp(-u) du, \beta \geq 0$
$\Psi(\beta)$	Digamma function, $\Psi(\beta) = \frac{d}{d\beta} \log \Gamma(\beta), \beta \geq 0$
$\mathcal{N}(\mathbf{u})$	Standard Gaussian density for $\mathbf{u} \in R^d$, as $\mathcal{N}(\mathbf{u}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\mathbf{u}'\mathbf{u}\right)$
$p_{\gamma^{-1}}(z; \alpha, \beta)$	Inverse Gamma distribution, $p_{\gamma^{-1}}(z) = \frac{\alpha^\beta}{2^\beta \Gamma(\beta)} z^{-\beta-1} \exp\left(-\frac{\alpha}{2z}\right), \beta > 1, \alpha > 0$
$H(\mathbf{x})$	Differential entropy of $\mathbf{x}$, $H(\mathbf{x}) = -\int_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}$
$I(\mathbf{x})$	Multi-information [7] of $\mathbf{x}$, $I(\mathbf{x}) = \int_{\mathbf{x}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\prod_{k=1}^d p(x_k)} d\mathbf{x}$

3 Properties of Gamma Function and Inverse Gamma Density

Fact 1 *([2]) $\int_0^\infty z^{-b-1} \exp\left(-\frac{a}{z}\right) dz = a^{-b}\Gamma(b)$ for $a > 0$ and $b > 1$.*

Proof (Fact 1): First, we change variable $u = a/z$ to have $z = a/u$ and $dz = -a/u^2 du$. Replacing z with u in $\int_0^\infty z^{-b-1} \exp\left(-\frac{a}{z}\right) dz$, we have

$$\int_0^\infty z^{-b-1} \exp\left(-\frac{a}{z}\right) dz = -\int_\infty^0 a^{-b-1} u^{b+1} \exp(-u) \frac{a}{u^2} du = a^{-b} \int_0^\infty u^{b-1} \exp(-u) du.$$

Note the integral in the last step is $\Gamma(b)$, the result is proved.

Fact 2 *The mean and mode of the inverse Gamma density $p_{\gamma^{-1}}(z; \alpha, \beta)$ is $\frac{\alpha}{2(\beta-1)}$ and $\frac{\alpha}{2(\beta+1)}$, respectively. Furthermore, the mean of $1/z$ under the inverse Gamma density is $2\beta/\alpha$.*

Proof (Fact 2): We discuss each case individually.

1 The mean of z is given as

$$\int_0^\infty z p_{\gamma^{-1}}(z; \alpha, \beta) dz = \int_0^\infty z \cdot \frac{\alpha^\beta}{2^\beta \Gamma(\beta)} z^{-\beta-1} \exp\left(-\frac{\alpha}{2z}\right) dz = \frac{\alpha^\beta}{2^\beta \Gamma(\beta)} \int_0^\infty z^{-\beta} \exp\left(-\frac{\alpha}{2z}\right) dz.$$

Using Fact 1, the last integral equals to $(a/2)^{-(\beta-1)}\Gamma(\beta-1)$. Canceling terms, and using the property of Gamma function that $\Gamma(\beta)/\Gamma(\beta-1)=\beta-1$, we obtain that the mean is $\alpha/[2(\beta-1)]$.

2 The mode of z is $\operatorname{argmax}_z p_{\gamma^{-1}}(z;\alpha,\beta)=\operatorname{argmax}_z \log p_{\gamma^{-1}}(z;\alpha,\beta)$. We have

$$\log p_{\gamma^{-1}}(z;\alpha,\beta)=\log\alpha^\beta-\log 2^\beta\Gamma(\beta)-(\beta+1)\log z-\frac{\alpha}{2z},$$

whose derivative with regards to z is $-\frac{\beta+1}{z}+\frac{\alpha}{2z^2}$. Setting the derivative to zero and solving for z, we obtain that the mode is $\alpha/[2(\beta+1)]$.

3 The mean of $1/z$ is given as

$$\int_0^\infty z^{-1}p_{\gamma^{-1}}(z;\alpha,\beta)dz=\frac{\alpha^\beta}{2^\beta\Gamma(\beta)}\int_0^\infty z^{-\beta-2}\exp\left(-\frac{\alpha}{2z}\right)dz.$$

Using Fact 1, the last integral equals to $(a/2)^{-(\beta+1)}\Gamma(\beta+1)$. Canceling terms, and using the property of Gamma function that $\Gamma(\beta+1)/\Gamma(\beta)=\beta$, we obtain that the mean of $1/z$ is $2\beta/\alpha$.

4 Properties of Multivariate *t* Model

Fact 3 *([3, 5]) The isotropic multivariate* t *model is a Gaussian scale mixture formed by an inverse Gamma mixing density and a standard Gaussian density, as:*

$$p_t(\mathbf{x};\alpha,\beta)\equiv\frac{\alpha^\beta\Gamma\left(\beta+d/2\right)}{\pi^{d/2}\Gamma(\beta)}(\alpha+\mathbf{x}'\mathbf{x})^{-\beta-d/2}=\int_0^\infty\mathcal{N}(\mathbf{x}/\sqrt{z})p_{\gamma^{-1}}(z;\alpha,\beta)dz.$$

Proof (Fact 3): First, we expand the joint dentsity of $\mathbf{x}$ and z as

$$\begin{aligned}\mathcal{N}(\mathbf{x}/\sqrt{z})p_{\gamma^{-1}}(z;\alpha,\beta) &= \frac{1}{(2\pi)^{d/2}z^{d/2}}\exp\left(-\frac{1}{2z}\mathbf{x}'\mathbf{x}\right)\frac{\alpha^\beta}{2^\beta\Gamma(\beta)}z^{-\beta-1}\exp\left(-\frac{\alpha}{2z}\right)\\ &= \frac{\alpha^\beta}{(2\pi)^{d/2}2^\beta\Gamma(\beta)}z^{-\beta-d/2-1}\exp\left(-\frac{1}{2z}\left(\alpha+\mathbf{x}'\mathbf{x}\right)\right).\end{aligned}$$

Using Fact 1, with $b=\beta+d/2$ and $a=\frac{1}{2}\left(\alpha+\mathbf{x}'\mathbf{x}\right)$, we have

$$\int_0^\infty z^{-\beta-d/2-1}\exp\left(-\frac{1}{2z}\left(\alpha+\mathbf{x}'\mathbf{x}\right)\right)dz=\left(\frac{1}{2}\left(\alpha+\mathbf{x}'\mathbf{x}\right)\right)^{-\beta-d/2}\Gamma(\beta+d/2).$$

Putting all terms together, we have

$$\begin{aligned}\int_0^\infty\mathcal{N}(\mathbf{x}/\sqrt{z})p_{\gamma^{-1}}(z;\alpha,\beta)dz &= \frac{\alpha^\beta}{(2\pi)^{d/2}2^\beta\Gamma(\beta)}\left(\frac{1}{2}\left(\alpha+\mathbf{x}'\mathbf{x}\right)\right)^{-\beta-d/2}\Gamma(\beta+d/2)\\ &= \frac{\alpha^\beta\Gamma\left(\beta+d/2\right)}{\pi^{d/2}\Gamma(\beta)}\frac{1}{(\alpha+\mathbf{x}'\mathbf{x})^{\beta+d/2}}.\end{aligned}$$

Note that Fact 3 can be extended to multivariate *t* models with non-diagonal covariance matrix by the simple technique of "unwhitening" that restore the covariance matrix.

Definition 1 *([6]) A density $p(\mathbf{x})$ is known as* re-normalizable *if any marginal density of a subset of vector $\mathbf{x}$ has the same parametric form (though may be with different parameters) as $p(\mathbf{x})$.*

Gaussian density is an example of re-normalizable density. Indeed, isotropic Gaussian is the only density that is re-normalizable, factorizable into mutually independent components, and rotation invariant at the same time, a fact known by Maxwell in the 1870s!

Fact 4 *Multivariate* t *models are re-normalizable. Specifically, the 1D marginal density for x_i of a d-dimensional* t *model is a 1D* t *model as:*

$$p_t(x_i;\alpha,\beta)=\frac{\alpha^\beta\Gamma\left(\beta+1/2\right)}{\pi^{1/2}\Gamma(\beta)}\frac{1}{(\alpha+\mathbf{x}'\mathbf{x})^{\beta+1/2}}.$$

Proof (Fact 4): First, we write out the marginal density of x_i according to its definition as:

$$p_t(x_i;\alpha,\beta) = \int_{\mathbf{x}_{\setminus i}} p_t(\mathbf{x};\alpha,\beta)d\mathbf{x}_{\setminus i} = \frac{\alpha^\beta\Gamma(\beta+d/2)}{\pi^{d/2}\Gamma(\beta)}\int_{\mathbf{x}_{\setminus i}}\left((\alpha+x_i^2)+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i}\right)^{-\beta-d/2}d\mathbf{x}_{\setminus i}$$

Note that the term inside the integral is the unnormalized *t* model for $(d-1)$-dimensional data $\mathbf{x}_{\setminus i}$, with $\alpha' = \alpha + x_i^2$, and $\beta' = \beta + 1/2$, whose normalizing property states that

$$\frac{(\alpha+x_i^2)^{\beta+1/2}\Gamma(\beta+d/2)}{\pi^{(d-1)/2}\Gamma(\beta+1/2)}\int_{\mathbf{x}_{\setminus i}}\left((\alpha+x_i^2)+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i}\right)^{-\beta-d/2}d\mathbf{x}_{\setminus i} = 1,$$

using which we have

$$\frac{\alpha^\beta\Gamma(\beta+d/2)}{\pi^{d/2}\Gamma(\beta)}\int_{\mathbf{x}_{\setminus i}}\left((\alpha+x_i^2)+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i}\right)^{-\beta-d/2}d\mathbf{x}_{\setminus i} = \frac{\alpha^\beta\Gamma(\beta+d/2)}{\pi^{d/2}\Gamma(\beta)}\frac{\pi^{(d-1)/2}\Gamma(\beta+1/2)}{(\alpha+x_i^2)^{\beta+1/2}\Gamma(\beta+d/2)}.$$

Rearranging terms in the above identity proves the result.

Lemma 1 *For a d-dimensional isotropic* t *vector* $\mathbf{x}$*, we have*

$$\mathrm{E}(x_i|\mathbf{x}_{\setminus i}) = 0, \;\; \textit{and} \;\; \mathrm{v}ar(x_i|\mathbf{x}_{\setminus i}) = \frac{1}{2\beta+d-3}(\alpha+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i}),$$

where $\mathrm{E}(\cdot)$ *and* $\mathrm{v}ar(\cdot)$ *denote expectation and variance, respectively.*

Proof (Lemma 1): We show the result in two steps.

1 First, for the conditional mean, using the GSM equivalence given in Fact 3, we have

$$E(x_i|\mathbf{x}_{\setminus i}) = \int_{x_i}\frac{x_i p_t(\mathbf{x};\alpha,\beta)}{p(\mathbf{x}_{\setminus i})}dx_i = \int_z\frac{p_{\gamma^{-1}}(z;\alpha,\beta)}{p(\mathbf{x}_{\setminus i})}dz\int_{x_i}x_i\mathcal{N}(\mathbf{x}/\sqrt{z})dx_i = 0,$$

the last step is due to that $\int_{x_i}x_i\mathcal{N}(\mathbf{x}/\sqrt{z})dx_i = 0$.

2 For the conditional variance, we have $\mathrm{v}ar(x_i|\mathbf{x}_{\setminus i}) = E(x_i^2|\mathbf{x}_{\setminus i}) - [E(x_i|\mathbf{x}_{\setminus i})]^2 = E(x_i^2|\mathbf{x}_{\setminus i})$. Hence, we need to compute

$$E(x_i^2|\mathbf{x}_{\setminus i}) = \frac{\int_{x_i}x_i^2p_t(\mathbf{x};\alpha,\beta)dx_i}{\int_{x_i}p_t(\mathbf{x};\alpha,\beta)dx_i} = \frac{\int_z p_{\gamma^{-1}}(z;\alpha,\beta)dz\int_{x_i}x_i^2\mathcal{N}(\mathbf{x}/\sqrt{z})dx_i}{\int_z p_{\gamma^{-1}}(z;\alpha,\beta)dz\int_{x_i}\mathcal{N}(\mathbf{x}/\sqrt{z})dx_i}.$$

Next, using the property of a Gaussian density, we have $\int_{x_i}x_i^2\mathcal{N}(\mathbf{x}/\sqrt{z})dx_i = z\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})$ and $\int_{x_i}\mathcal{N}(\mathbf{x}/\sqrt{z})dx_i = \mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})$. Substituting these results into that of $E(x_i^2|\mathbf{x}_{\setminus i})$, we have

$$E(x_i^2|\mathbf{x}_{\setminus i}) = \frac{\int_z zp_{\gamma^{-1}}(z;\alpha,\beta)\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})dz}{\int_z p_{\gamma^{-1}}(z;\alpha,\beta)\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})dz}.$$

Next, we use the recursive definition of the inverse Gamma density $zp_{\gamma^{-1}}(z;\alpha,\beta-1) = \frac{\alpha\Gamma(\beta-1)}{2\Gamma(\beta)}p_{\gamma^{-1}}(z;\alpha,\beta-1)$ to obtain

$$E(x_i^2|\mathbf{x}_{\setminus i}) = \frac{\alpha\Gamma(\beta-1)}{2\Gamma(\beta)}\frac{\int_z p_{\gamma^{-1}}(z;\alpha,\beta-1)\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})dz}{\int_z p_{\gamma^{-1}}(z;\alpha,\beta)\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})dz},$$

which is further simplified to

$$E(x_i^2|\mathbf{x}_{\setminus i}) = \frac{\alpha}{2(\beta-1)}\frac{\int_z p_{\gamma^{-1}}(z;\alpha,\beta-1)\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})dz}{\int_z p_{\gamma^{-1}}(z;\alpha,\beta)dz\mathcal{N}(\mathbf{x}_{\setminus i}/\sqrt{z})dz}.$$

Using the GSM equivalency of *t* model given in Fact 3, we see that the numerator is a *t* model of $d-1$ dimension with parameter α and $\beta-1$, and the denominator is a *t* model of $d-1$ dimension with parameter α and β. Writing the density explicitly, we have

$$E(x_i^2|\mathbf{x}_{\setminus i}) = \frac{\alpha}{2(\beta-1)}\frac{\frac{\alpha^{\beta-1}\Gamma(\beta-1+(d-1)/2)}{\pi^{(d-1)/2}\Gamma(\beta-1)}(\alpha+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i})^{-(\beta-1)-(d-1)/2}}{\frac{\alpha^{\beta}\Gamma(\beta+(d-1)/2)}{\pi^{(d-1)/2}\Gamma(\beta)}(\alpha+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i})^{-\beta-(d-1)/2}}.$$

After simplification, we obtain $E(x_i^2|\mathbf{x}_{\setminus i}) = \frac{1}{2\beta+d-3}(\alpha+\mathbf{x}'_{\setminus i}\mathbf{x}_{\setminus i})$.

Lemma 2 *For the d-dimensional isotropic* t *vector* $\mathbf{x}$ *with parameters* (α, β)*, we consider three estimators of z as: (i) the MAP estimator,* $\hat{z}_1 = \mathrm{argmax}_z\, p(z|\mathbf{x})$*, which is the mode of the posterior density, (ii) the BLS estimator, which is the mean of the posterior density* $\hat{z}_2 = \mathrm{E}_{z|\mathbf{x}}(z|\mathbf{x})$*, and (iii) the inverse of the BLS estimator of* $1/z$*, as* $\hat{z}_3 = \left(\mathrm{E}_{z|\mathbf{x}}\left(1/z|\mathbf{x}\right)\right)^{-1}$*. We can express them as*

$$\hat{z}_1 = \frac{\alpha + \mathbf{x}'\mathbf{x}}{2\beta + d + 2}, \qquad \hat{z}_2 = \frac{\alpha + \mathbf{x}'\mathbf{x}}{2\beta + d - 2}, \text{ and } \hat{z}_3 = \left(E_{z|\mathbf{x}}\left(1/z|\mathbf{x}\right)\right)^{-1} = \frac{\alpha + \mathbf{x}'\mathbf{x}}{2\beta + d}.$$

Proof (Lemma 2):

$$p(z|\mathbf{x};\alpha,\beta) = \frac{p(\mathbf{x},z;\alpha,\beta)}{p_t(\mathbf{x};\alpha,\beta)} = \frac{\frac{1}{(2\pi)^{d/2}z^{d/2}}\exp\left(-\frac{1}{2z}\mathbf{x}'\mathbf{x}\right)\frac{\alpha^\beta}{2^\beta\Gamma(\beta)}z^{-\beta-1}\exp\left(-\frac{\alpha}{2z}\right)}{\frac{\alpha^\beta\Gamma(\beta+d/2)}{\pi^{d/2}\Gamma(\beta)}\frac{1}{(\alpha+\mathbf{x}'\mathbf{x})^{\beta+d/2}}}.$$

Rearranging terms, we have

$$p(z|\mathbf{x};\alpha,\beta) = \frac{(\alpha+\mathbf{x}'\mathbf{x})^{\beta+d/2}}{2^{\beta+d/2}\Gamma\left(\beta+d/2\right)}z^{-\beta-d/2-1}\exp\left(-\frac{1}{2z}\left(\alpha+\mathbf{x}'\mathbf{x}\right)\right),$$

which is an inverse Gamma density with parameters $\alpha' = \alpha + \mathbf{x}'\mathbf{x}$, and $\beta' = \beta + d/2$.

Now using Fact 2, we have

1 The MAP estimator is the mode of the posterior inverse Gamma density and is then

$$\hat{z}_1 = \frac{\alpha'}{2(\beta'+1)} = \frac{\alpha+\mathbf{x}'\mathbf{x}}{2\beta+d+2}.$$

2 The BLS estimator is the mean of the posterior inverse Gamma density and is then

$$\hat{z}_2 = \frac{\alpha'}{2(\beta'-1)} = \frac{\alpha+\mathbf{x}'\mathbf{x}}{2\beta+d-2}.$$

3 The mean of $1/z$ with regards to the posterior inverse Gamma density and is

$$\frac{2(\beta')}{\alpha'} = \frac{2\beta+d}{\alpha+\mathbf{x}'\mathbf{x}}.$$

Therefore, we have $\hat{z}_3 = \frac{\alpha+\mathbf{x}'\mathbf{x}}{2\beta+d}$.

5 Maximum Likelihood Fitting of *t* Model

Here we describe the procedure to fit the *t* model to data using maximum likelihood. To simplify computation, we assume the training data have been centered and whitened, and thus they have zero mean and their covariance matrix is an identity matrix. Enforcing this constraint on the *t* model, we have

$$I = \int_{\mathbf{x}} \mathbf{x}\mathbf{x}' p_t(\mathbf{x};\alpha,\beta)d\mathbf{x} = \int_z p_{\gamma^{-1}}(z;\alpha,\beta)dz \int_{\mathbf{x}} \mathbf{x}\mathbf{x}'\mathcal{N}(\mathbf{x}/\sqrt{z})d\mathbf{x} = I\int_z z p_{\gamma^{-1}}(z;\alpha,\beta)dz.$$

With Fact 2, we have that $\int_z z p_{\gamma^{-1}}(z;\alpha,\beta)dz = \frac{\alpha}{2(\beta-1)}$. Therefore, this leads to $\alpha = 2(\beta-1)$.

Next, β is estimated from training data by maximizing likelihood. Specifically, with α removed from the model, the average log likelihood function for the *t* model is

$$\begin{aligned} L(\beta) = \frac{1}{N}\sum_{n=1}^N \log p_t(\mathbf{x}_n; 2(\beta-1),\beta) &= \beta\log 2(\beta-1) + \log\Gamma(\beta+d/2) - \frac{d}{2}\log\pi \\ &- \log\Gamma(\beta) - \frac{\beta+d/2}{N}\sum_{n=1}^N \log(2(\beta-1)+\mathbf{x}_n'\mathbf{x}_n). \end{aligned}$$

Optimal β is the root of the nonlinear equation $\frac{d}{d\beta}L(\beta) = 0$, which is obtained by a numerical Newton-Raphson procedure.

6 Properties of the DN Transform

We adopt the standard form of the divisive normalization (DN) transform in the main submission as

$$\phi(\mathbf{x}) \equiv \frac{\mathbf{x}}{\sqrt{\alpha + \mathbf{x}'\mathbf{x}}} = \frac{\|\mathbf{x}\|}{\sqrt{\alpha + \|\mathbf{x}\|^2}} \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{1}$$

Lemma 3 *For the standard DN transform given in Eq. (1), its inversion for* $\mathbf{y} \in R^d$ *with* $\|\mathbf{y}\| < 1$ *is* $\phi^{-1}(\mathbf{y}) = \frac{\sqrt{\alpha}\mathbf{y}}{\sqrt{1-\|\mathbf{y}\|^2}} = \frac{\sqrt{\alpha}\|\mathbf{y}\|}{\sqrt{1-\|\mathbf{y}\|^2}} \frac{\mathbf{y}}{\|\mathbf{y}\|}$. *The determinant of its Jacobian matrix is also in close form, which is given by*

$$\det\left(\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}}\right) = \frac{\alpha}{(\alpha + \mathbf{x}'\mathbf{x})^{d/2+1}}.$$

Lemma 4 *If* $\mathbf{x} \in R^d$ *has an isotropic* t *density with parameter* (α, β)*, then its DN transform,* $\mathbf{y} = \phi(\mathbf{x})$*, follows an isotropic* τ *model [4], whose probability density function is*

$$p_\tau(\mathbf{y}) = \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} (1 - \mathbf{y}'\mathbf{y})_+^{\beta-1}, \tag{2}$$

where $(\cdot)_+$ *is the rectifying function.*

Proof (Lemma 4): First, note that

$$\mathbf{y} = \frac{\mathbf{x}}{\sqrt{\alpha + \mathbf{x}'\mathbf{x}}} \Rightarrow \mathbf{y}'\mathbf{y} = \frac{\mathbf{x}'\mathbf{x}}{\alpha + \mathbf{x}'\mathbf{x}} \Rightarrow \|\mathbf{x}\| = \frac{\sqrt{\alpha}\|\mathbf{y}\|}{\sqrt{1 - \mathbf{y}'\mathbf{y}}}.$$

Further, as $\mathbf{x}/\|\mathbf{x}\| = \mathbf{y}/\|\mathbf{y}\|$, putting together, the inversion of the DN transform is given by $\phi^{-1}(\mathbf{y}) = \frac{\sqrt{\alpha}\mathbf{y}}{\sqrt{1-\|\mathbf{y}\|^2}} = \frac{\sqrt{\alpha}\|\mathbf{y}\|}{\sqrt{1-\|\mathbf{y}\|^2}} \frac{\mathbf{y}}{\|\mathbf{y}\|}$.

Next, we write the Jacobian matrix of the DN transform explicitly as:

$$J_\phi(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} \frac{\mathbf{x}}{\sqrt{\alpha + \mathbf{x}'\mathbf{x}}} = \frac{1}{\sqrt{\alpha + \mathbf{x}'\mathbf{x}}} I_d - \frac{1}{\sqrt{\alpha + \mathbf{x}'\mathbf{x}}} \frac{\mathbf{x}\mathbf{x}'}{\alpha + \mathbf{x}'\mathbf{x}} = \frac{1}{\sqrt{\alpha + \mathbf{x}'\mathbf{x}}} \left(I_d - \frac{\mathbf{x}\mathbf{x}'}{\alpha + \mathbf{x}'\mathbf{x}}\right).$$

Therefore, we have

$$\det(J_\phi(\mathbf{x})) = (\alpha + \mathbf{x}'\mathbf{x})^{-d/2} \det\left(I_d - \frac{\mathbf{x}\mathbf{x}'}{\alpha + \mathbf{x}'\mathbf{x}}\right).$$

Now with identity $\det(I_d - \mathbf{v}\mathbf{v}') = (1 - \mathbf{v}'\mathbf{v})$ [1], we have

$$\det(J_\phi(\mathbf{x})) = (\alpha + \mathbf{x}'\mathbf{x})^{-d/2} \left(1 - \frac{\mathbf{x}'\mathbf{x}}{\alpha + \mathbf{x}'\mathbf{x}}\right) = \frac{\alpha}{(\alpha + \mathbf{x}'\mathbf{x})^{d/2+1}}.$$

Lemma 5 *If* $\mathbf{x} \in R^d$ *has an isotropic* t *density with parameter* (α, β)*, then its DN transform,* $\mathbf{y} = \phi(\mathbf{x})$*, follows an isotropic* τ *model [4], whose probability density function is*

$$p_\tau(\mathbf{y}) = \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} (1 - \mathbf{y}'\mathbf{y})_+^{\beta-1}, \tag{3}$$

where $(\cdot)_+$ *is the rectifying function.*

Proof (Lemma 5): Relation between densities of transformed random vector gives that $p(\mathbf{x}) = p(\phi(x)) \det(J_\phi(\mathbf{x}))$, or $p(\mathbf{y}) = p(\mathbf{x}) \frac{1}{\det(J_\phi(\mathbf{x}))}$. Using the Jacobian determinant of the DN transform given in Lemma 4, we have

$$p(\mathbf{y}) = \frac{\alpha^\beta \Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} \frac{1}{(\alpha + \mathbf{x}'\mathbf{x})^{\beta+d/2}} \times \alpha(\alpha + \mathbf{x}'\mathbf{x})^{d/2+1} = \frac{\alpha^{\beta-1}\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} \frac{1}{(\alpha + \mathbf{x}'\mathbf{x})^{\beta-1}}.$$

Next, with the inversion of the DN transform given in Lemma 4, we substitute $\mathbf{x} = \phi^{-1}(\mathbf{y})$ on the right hand side of the above identity to have

$$p(\mathbf{y}) = \frac{\alpha^{\beta-1}\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} \frac{1}{\left(\alpha + \frac{\alpha\mathbf{y}'\mathbf{y}}{(1-\mathbf{y}'\mathbf{y})_+}\right)^{\beta-1}} = \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} (1 - \mathbf{y}'\mathbf{y})_+^{\beta-1}.$$

7 Properties of Multivariate τ Model

Fact 5 *The multivariate τ models are re-normalizable. Specifically, the marginal density of y_i from a d dimensional τ with parameter β model is a 1D τ model with parameter $\beta + (d-1)/2$, as:*

$$p_\tau(y_i) = \frac{\Gamma(\beta + d/2)}{\pi^{1/2}\Gamma(\beta + (d-1)/2)} \left(1 - y_i^2\right)_+^{\beta+(d-3)/2}.$$

Proof (Fact 5): First using definition, we can write $p(y_i)$ explicitly as:

$$\begin{aligned} p_\tau(y_i) &= \int_{\mathbf{y}_{\setminus i}} p_\tau(\mathbf{y};\beta) d\mathbf{y}_{\setminus i} = \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} \int_{\mathbf{y}_{\setminus i}} \left((1-y_i^2) - \mathbf{y}'_{\setminus i}\mathbf{y}_{\setminus i}\right)_+^{\beta-1} d\mathbf{y}_{\setminus i} \\ &= \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} (1-y_i^2)_+^{\beta-1} \int_{\mathbf{y}_{\setminus i}} \left(1 - \frac{\mathbf{y}'_{\setminus i}\mathbf{y}_{\setminus i}}{1-y_i^2}\right)_+^{\beta-1} d\mathbf{y}_{\setminus i} \end{aligned}$$

Next introduce $d-1$ dimensional vector $\mathbf{u} = \frac{\mathbf{y}_{\setminus i}}{\sqrt{1-y_i^2}}$, we have

$$p_\tau(y_i) = \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} (1-y_i^2)_+^{\beta-1+(d-1)/2} \int_{\mathbf{u}} (1 - \mathbf{u}'\mathbf{u})_+^{\beta-1} d\mathbf{u}.$$

Note that the integral is an unnormalized $d-1$ dimensional τ vector with parameter β. Therefore we have

$$p_\tau(y_i) = \frac{\Gamma(\beta + d/2)}{\pi^{d/2}\Gamma(\beta)} (1-y_i^2)_+^{\beta-1+(d-1)/2} \frac{\pi^{(d-1)/2}\Gamma(\beta)}{\Gamma(\beta + (d-1)/2)}.$$

After we rearrange terms, the result follows.

8 Entropy and Multi-information of Multivariate *t* and τ Models

Fact 6 *The differential entropy of a d-dimensional* t *vector $\mathbf{x}$ is*

$$H(\mathbf{x}) = \frac{d}{2} \log \alpha\pi + \log\Gamma(\beta) - \log\Gamma(\beta + d/2) + (\beta + d/2)\left[\Psi(\beta + d/2) - \Psi(\beta)\right].$$

Similarly, the differential entropy of a d-dimensional τ vector $\mathbf{y}$ is

$$H(\mathbf{y}) = \frac{d}{2} \log \pi + \log\Gamma(\beta) - \log\Gamma(\beta + d/2) + (\beta - 1)\left[\Psi(\beta + d/2) - \Psi(\beta)\right].$$

Proof (Fact 6): We prove each case separately.

***t* model**. Using the density fuction of *t* model and the definition of differential entropy, we have

$$\begin{aligned} H(\mathbf{x}) &= \frac{d}{2}\log\pi - \beta\log\alpha + \log\Gamma(\beta) - \log\Gamma(\beta + d/2) \\ &+ (\beta + d/2)\int_{\mathbf{x}} \log(\alpha + \mathbf{x}'\mathbf{x}) p_t(\mathbf{x};\alpha,\beta) d\mathbf{x}. \end{aligned}$$

All we need to compute is the last integral, and we next show that

$$\int_{\mathbf{x}} \log(\alpha + \mathbf{x}'\mathbf{x}) p_t(\mathbf{x};\alpha,\beta) d\mathbf{x} = \log\alpha + \Psi(\beta + d/2) - \Psi(\beta).$$

First, by the normalizing property of the *t* model density function, we have

$$\int_{\mathbf{x}} (\alpha + \mathbf{x}'\mathbf{x})^{-\beta-d/2} d\mathbf{x} = \frac{\pi^{d/2}\Gamma(\beta)}{\alpha^\beta\Gamma(\beta + d/2)}.$$

Next, take derivative w.r.t β to both sides, we have

$$-\int_{\mathbf{x}} (\alpha + \mathbf{x}'\mathbf{x})^{-\beta-d/2} \log(\alpha + \mathbf{x}'\mathbf{x}) d\mathbf{x} = \frac{\partial}{\partial\beta}\left(\frac{\pi^{d/2}\Gamma(\beta)}{\alpha^\beta\Gamma(\beta + d/2)}\right),$$

then multiply $-\alpha^{\beta}\Gamma(\beta+d/2)/\pi^{d/2}\Gamma(\beta)$ on both sides to obtain

$$\begin{aligned}\int_{\mathbf{x}} \log(\alpha+\mathbf{x}'\mathbf{x})p_t(\mathbf{x};\alpha,\beta)d\mathbf{x} &= -\frac{\partial}{\partial\beta}\left(\frac{\pi^{d/2}\Gamma(\beta)}{\alpha^{\beta}\Gamma(\beta+d/2)}\right)\frac{\alpha^{\beta}\Gamma(\beta+d/2)}{\pi^{d/2}\Gamma(\beta)} \\ &= \frac{\partial}{\partial\beta}\log\frac{\alpha^{\beta}\Gamma(\beta+d/2)}{\Gamma(\beta)} = \log\alpha+\Psi(\beta+d/2)-\Psi(\beta).\end{aligned}$$

Last, substitute the result to the entropy computation will prove the result.

τ **model.** First, using the normalization property of the τ model, we have

$$\frac{\pi^{d/2}\Gamma(\beta)}{\Gamma(\beta+d/2)} = \int_{\mathbf{y}}(1-\mathbf{y}'\mathbf{y})_+^{\beta-1}\,d\mathbf{y}.$$

Next, taking derivatives with regards to β on both sides yields

$$\pi^{d/2}\frac{d}{d\beta}\frac{\Gamma(\beta)}{\Gamma(\beta+d/2)} = \frac{d}{d\beta}\int_{\mathbf{y}}(1-\mathbf{y}'\mathbf{y})_+^{\beta-1}\,d\mathbf{y} = \int_{\mathbf{y}}(1-\mathbf{y}'\mathbf{y})_+^{\beta-1}\log(1-\mathbf{y}'\mathbf{y})_+\,d\mathbf{y}.$$

Then multiplying $\frac{\Gamma(\beta+d/2)}{\pi^{d/2}\Gamma(\beta)}$ on both sides further yields:

$$\frac{d}{d\beta}\log\frac{\Gamma(\beta)}{\Gamma(\beta+d/2)} = \Psi(\beta)-\Psi(\beta+d/2) = \int_{\mathbf{y}}p_\tau(\mathbf{y};\beta)\log(1-\mathbf{y}'\mathbf{y})_+\,d\mathbf{y}. \tag{4}$$

Next, logarithm of the τ model is

$$\log p_\tau(\mathbf{y};\beta) = \log\Gamma(\beta+d/2)-\log\Gamma(\beta)-d/2\log\pi+(\beta-1)\log(1-\mathbf{y}'\mathbf{y})_+,$$

therefore the entropy is

$$H(\mathbf{y}) = d/2\log\pi+\log\Gamma(\beta)-\log\Gamma(\beta+d/2)-(\beta-1)\int_{\mathbf{y}}p_\tau(\mathbf{y};\beta)\log(1-\mathbf{y}'\mathbf{y})_+\,d\mathbf{y}.$$

Next, using Eq.(4) to replace the integral in the last term, the result follows.

Lemma 6 *The MI of a d-dimensional isotropic* t *model ($\Sigma = I$) is*

$$\begin{aligned}I(\mathbf{x}) &= (d-1)\log\Gamma(\beta)-d\log\Gamma\left(\beta+\frac{1}{2}\right)+\log\Gamma\left(\beta+\frac{d}{2}\right) \\ &+ d\left(\beta+\frac{1}{2}\right)\Psi\left(\beta+\frac{1}{2}\right)-\left(\beta+\frac{d}{2}\right)\Psi\left(\beta+\frac{d}{2}\right)-(d-1)\beta\Psi(\beta).\end{aligned}$$

The multi-information of a d-dimensional τ model with density in Eq.(3) is

$$\begin{aligned}I(\mathbf{y}) &= d\log\Gamma\left(\beta+\frac{d-1}{2}\right)-\log\Gamma(\beta)-(d-1)\log\Gamma\left(\beta+\frac{d}{2}\right)+(\beta-1)\Psi(\beta) \\ &+ (d-1)\left(\beta+\frac{d}{2}-1\right)\Psi\left(\beta+\frac{d}{2}\right)-d\left(\beta+\frac{d-3}{2}\right)\Psi\left(\beta+\frac{d-1}{2}\right).\end{aligned}$$

In both cases, $\Psi(\beta)$ denotes the Digamma function which is defined as $\Psi(\beta)=\frac{d}{d\beta}\log\Gamma(\beta)$.

Proof (Lemma 6): To compute multi-information, we use its relation with the differential entropy, as $I(\mathbf{x}) = \sum_{k=1}^{d}H(x_k)-H(\mathbf{x})$ and $I(\mathbf{x}) = \sum_{k=1}^{d}H(y_k)-H(\mathbf{y})$. $H(\mathbf{x})$ and $H(\mathbf{y})$ can be directly obtained using Fact 6. On the other hand, as we consider only isotropic models, we have $H(x_k)=H(x'_k)$ and $H(y_k)=H(y'_k)$ for $k\neq k'$.

To compute $H(x_k)$, we make use of Fact 4 that states *t* model is re-normalizable, so we can simply use the parameters from the 1D marginal *t* density in the entropy given in Fact 6 to obtain,

$$H(x_i) = \frac{1}{2}\log\alpha\pi+\log\Gamma(\beta)-\log\Gamma(\beta+1/2)+(\beta+1/2)\left[\Psi(\beta+1/2)-\Psi(\beta)\right].$$

Similarly, $H(y_k)$ can be obtained using the re-normalizability of τ model as

$$\begin{aligned}H(y_i) &= \frac{1}{2}\log\pi+\log\Gamma\left(\beta+\frac{d-1}{2}\right)-\log\Gamma\left(\beta+\frac{d}{2}\right) \\ &+ \left(\beta+\frac{d-3}{2}\right)\left[\Psi\left(\beta+\frac{d}{2}\right)-\Psi\left(\beta+\frac{d-1}{2}\right)\right].\end{aligned}$$

Putting all these results together and after some tedious algebraical manipulations, we obtain $I(\mathbf{x})$ and $I(\mathbf{y})$ as stated in the Lemma.

References

[1] K. M. Abadir and J. R. Magnus. *Matrix Algebra*. Cambridge, 2005. 5

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964. 1

[3] D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1):99–102, 1974. 2

[4] J. Costa, A. Hero, and C. Vignat. On solutions to multivariate maximum α-entropy problems. In *EMM-CVPR*, 2003. 5

[5] Samuel Kotz and Saralees Nadarajah. *Multivariate t Distributions and Their Applications*. Cambridge University Press, 2004. 2

[6] Renio S. Mendes and Constantino Tsallis. Renormalization group approach to nonextensive statistical mechanics. *Physics Letters A*, 285:273–278, July 2001. 2

[7] M. Studeny and J. Vejnarova. The multiinformation function as a tool for measuring stochastic dependence. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 261–297. Dordrecht: Kluwer., 1998. 1