

## FAST TRACK COMMUNICATION

**A closed-form expression for the Sharma–Mittal entropy of exponential families**Frank Nielsen<sup>1,2</sup> and Richard Nock<sup>3</sup><sup>1</sup> Sony Computer Science Laboratories, Inc. Tokyo, Japan<sup>2</sup> École Polytechnique, CS Department LIX, Palaiseau, France<sup>3</sup> UAG-CEREGMIA, Martinique, FranceE-mail: [Frank.Nielsen@acm.org](mailto:Frank.Nielsen@acm.org) and [Richard.Nock@martinique.univ-ag.fr](mailto:Richard.Nock@martinique.univ-ag.fr)

Received 31 August 2011, in final form 4 November 2011

Published DD MMM 2011

Online at [stacks.iop.org/JPhysA/45/000000](http://stacks.iop.org/JPhysA/45/000000)**Abstract**

The Sharma–Mittal entropies generalize the celebrated Shannon, Rényi and Tsallis entropies. We report a closed-form formula for the Sharma–Mittal entropies and relative entropies for arbitrary exponential family distributions. We explicitly instantiate the formula for the case of the multivariate Gaussian distributions and discuss on its estimation.

Q1

PACS numbers: 03.65.Ta, 03.67.–a

(Some figures may appear in colour only in the online journal)

Q2

The Sharma–Mittal entropy  $H_{\alpha,\beta}(p)$  [1, 2] of a probability density<sup>4</sup>  $p$  is defined as

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} \left( \left( \int p(x)^\alpha dx \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \text{ with } \alpha > 0, \alpha \neq 1 \text{ and } \beta \neq 1. \quad (1)$$

This bi-parametric family of entropies tends in limit cases to Rényi entropies  $R_\alpha(p) = \frac{1}{1-\alpha} \log \int p(x)^\alpha dx$  (for  $\beta \rightarrow 1$ ), Tsallis entropies  $T_\alpha(p) = \frac{1}{1-\alpha} (\int p(x)^\alpha dx - 1)$  (for  $\beta \rightarrow \alpha$ ) and Shannon entropy  $H(p) = - \int p(x) \log p(x) dx$  (for both  $\alpha, \beta \rightarrow 1$ ). The Sharma–Mittal entropy has previously been studied in the context of multi-dimensional harmonic oscillator systems [3].

Many usual statistical distributions including the Gaussians and discrete multinomials (that is, normalized histograms) belong to the exponential families [4]. Those exponential families play a major role in the field of thermo-statistics [5] and admit the generic canonical decomposition

$$p_F(x|\theta) = \exp(\langle \theta, t(x) \rangle - F(\theta) + k(x)), \quad (2)$$

<sup>4</sup> For the sake of simplicity and without loss of generality, we consider the probability density function  $p$  of a continuous random variable  $X \sim p$  in this note. For multivariate densities  $p$ , the integral notation  $\int$  denotes the corresponding multi-dimensional integral, so that we write for short  $\int p(x) dx = 1$ . Our results hold for probability mass functions and probability measures in general.

where  $\langle \cdot, \cdot \rangle$  denotes the inner product,  $F$  is a strictly convex  $C^\infty$  function characterizing the family (called the log normalizer since  $F(\theta) = \log \int e^{\langle \theta, t(x) \rangle + k(x)} dx$ ),  $\theta \in \Theta$  is the natural parameter denoting the member of the family  $\mathcal{E}_F = \{p_F(x|\theta) \mid \theta \in \Theta\}$ ,  $t(x)$  is the sufficient statistics and  $k(x)$  is an auxiliary carrier measure [4]. The natural parameter space  $\Theta = \{\theta \mid p_F(x; \theta) < \infty\}$  is an open convex set.

For example, the probability density of a multivariate Gaussian  $p \sim N(\mu, \Sigma)$  centered at  $\mu$  with a positive-definite covariance matrix  $\Sigma$  is conventionally written as

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|\Sigma|}} \exp - \frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2}, \tag{3}$$

where  $|\Sigma| > 0$  denotes the determinant of the positive-definite matrix. Rewriting the density of equation (3) to fit the canonical decomposition of equation (2), we obtain

$$p(x|\mu, \Sigma) = \exp \left( -\frac{1}{2} x^T \Sigma^{-1} x + x^T \Sigma^{-1} \mu - \frac{1}{2} \mu^T \Sigma^{-1} \mu - \frac{1}{2} \log(2\pi)^d |\Sigma| \right). \tag{4}$$

Using the matrix trace cyclic property, we have  $-\frac{1}{2} x^T \Sigma^{-1} x = \text{tr}(-\frac{1}{2} x^T \Sigma^{-1} x) = \text{tr}(x x^T \times (-\frac{1}{2} \Sigma^{-1}))$ , where  $\text{tr}$  denotes the matrix trace operator. It follows that

$$p(x|\mu, \Sigma) = \exp \left( \langle (x, x x^T), (\Sigma^{-1} \mu, -\frac{1}{2} \Sigma^{-1}) \rangle - F(\theta) \right), \tag{5}$$

$$= p(x|\theta), \tag{6}$$

with  $\theta = (\Sigma^{-1} \mu, -\frac{1}{2} \Sigma^{-1})$  and  $F(\theta) = \frac{1}{2} \log(2\pi)^d |\Sigma| + \frac{1}{2} \mu^T \Sigma^{-1} \mu$  (and  $k(x) = 0$ ). In this decomposition, the natural parameter  $\theta = (\Sigma^{-1} \mu, -\frac{1}{2} \Sigma^{-1}) = (v, M)$  consists of two parts: a vectorial part  $v$  and a symmetric negative definite matrix part  $M < 0$ . The inner product of  $\theta = (v, M)$  and  $\theta' = (v', M')$  is defined as  $\langle \theta, \theta' \rangle = v^T v' + \text{tr}(M^T M')$ . For univariate normal distributions, the natural parameter  $\theta$  is  $(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$ . The order of the exponential family is the dimension of its natural parameter space  $\Theta$ . Normal  $d$ -dimensional distributions  $N(\mu, \Sigma)$  form an exponential family of order  $d + \frac{d(d+1)}{2} = \frac{d(d+3)}{2}$ .

We have  $M = -\frac{1}{2} \Sigma^{-1}$ , that is,  $|\Sigma^{-1}| = |\Sigma|^{-1} = |-2M|$  and  $\mu^T = -\frac{1}{2} v^T M^{-1}$  (since  $M^{-1} = -2\Sigma$ ,  $-\frac{1}{2} M^{-1} v = \Sigma v = \mu$  and  $M^{-T} = M^{-1}$ ). It follows that the log normalizer  $F$  expressed using the canonical natural parameters is

$$F(\mu, \Sigma) = \frac{1}{2} \log(2\pi)^d |\Sigma| + \frac{1}{2} \mu^T \Sigma^{-1} \mu, \tag{7}$$

$$F(v, M) = \frac{d}{2} \log 2\pi - \frac{1}{2} \log |-2M| - \frac{1}{4} v^T M^{-1} v. \tag{8}$$

In order to calculate the Sharma–Mittal entropy of equation (1), let  $M_\alpha(p) = \int p(x)^\alpha dx$  so that

$$H_{\alpha, \beta}(p) = \frac{1}{1 - \beta} (M_\alpha(p)^{\frac{1-\beta}{1-\alpha}} - 1). \tag{9}$$

Let us prove that for an arbitrary exponential family  $\mathcal{E}_F = \{p_F(x|\theta) \mid \theta \in \Theta\}$ ,

$$M_\alpha(p) = e^{F(\alpha\theta) - \alpha F(\theta)} E_p[e^{(\alpha-1)k(x)}] \tag{10}$$

**Proof.**

$$M_\alpha(p) = \int e^{\alpha(t(x),\theta) - F(\theta) + k(x)} dx \tag{11}$$

$$= \int e^{(t(x),\alpha\theta) - \alpha F(\theta) + \alpha k(x) + (1-\alpha)k(x) - (1-\alpha)k(x) + F(\alpha\theta) - F(\alpha\theta)} dx \tag{12}$$

$$= \int e^{F(\alpha\theta) - \alpha F(\theta)} p_F(x; \alpha\theta) e^{(\alpha-1)k(x)} dx \tag{13}$$

$$= e^{F(\alpha\theta) - \alpha F(\theta)} \int p_F(x; \alpha\theta) e^{(\alpha-1)k(x)} dx \tag{14}$$

$$= e^{F(\alpha\theta) - \alpha F(\theta)} E_p[e^{(\alpha-1)k(x)}]. \tag{15}$$

Observe that in equation (13), we require  $\alpha\theta \in \Theta$  for a valid exponential family distribution. This is the case whenever the natural parameter space  $\Theta$  is a convex cone (e.g. Gaussian case). It follows from equation (10) that the Sharma–Mittal entropy of a distribution  $p \sim \mathcal{E}_F$  belonging to an exponential family  $\mathcal{E}_F$  is

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} (e^{\frac{1-\beta}{1-\alpha}(F(\alpha\theta) - \alpha F(\theta))} (E_p[e^{(\alpha-1)k(x)}])^{\frac{1-\beta}{1-\alpha}} - 1). \tag{16}$$

In particular, when the auxiliary carrier measure  $k(x) = 0$  [4] (including the above-mentioned multivariate Gaussian family), equation (16) becomes a closed-form formula since  $E_p[e^{(\alpha-1)k(x)}] = E_p[1] = 1$ :

$$H_{\alpha,\beta}(p) = \frac{1}{1-\beta} (e^{(1-\beta)\frac{F(\alpha\theta) - \alpha F(\theta)}{1-\alpha}} - 1), \tag{17}$$

$$= \frac{1}{1-\beta} ((e^{F(\alpha\theta) - \alpha F(\theta)})^{\frac{1-\beta}{1-\alpha}} - 1). \tag{18}$$

We derive in limit cases expressions for the Rényi, Tsallis and Shannon entropies of an arbitrary exponential family (with  $k(x) = 0$ ):

$$R_\alpha(p) = \lim_{\beta \rightarrow 1} H_{\alpha,\beta}(p) = \frac{1}{1-\alpha} (F(\alpha\theta) - \alpha F(\theta)), \tag{19}$$

$$T_\alpha(p) = \lim_{\beta \rightarrow \alpha} H_{\alpha,\beta}(p) = \frac{1}{1-\alpha} (e^{F(\alpha\theta) - \alpha F(\theta)} - 1), \tag{20}$$

$$H(p) = \lim_{\beta, \alpha \rightarrow 1} H_{\alpha,\beta}(p) = F(\theta) - \langle \theta, \nabla F(\theta) \rangle. \tag{21}$$

Note that the Shannon entropy of a member of an exponential family  $p \sim \mathcal{E}_F$  indexed with a natural parameter  $\theta$  can also be rewritten as  $H(p) = H(\theta) = -F^*(\eta)$ , with  $\eta = \nabla F(\theta)$  being the dual moment coordinates and  $F^*$  the Legendre  $C^\infty$  convex conjugate of  $F$  [6].

Let us instantiate the generic formula of equation (17) to the case of multivariate Gaussians with the mean parameter  $\mu$  and covariance matrix  $\Sigma$ . We obtain

$$H_{\alpha,\beta}(\mu, \Sigma) = \frac{1}{1-\beta} \left( \frac{((2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}})^{1-\beta}}{\alpha^{\frac{d(1-\beta)}{2(1-\alpha)}}} - 1 \right) \tag{22}$$

independent of  $\mu$  (in 1D,  $\Sigma = \sigma^2$  so that  $|\Sigma|^{\frac{1}{2}} = \sigma$ ). Indeed, consider the expression  $F(\alpha\theta) - \alpha F(\theta)$  in equation (16) for the Gaussian log normalizer of equation (8). Using the fact that  $|\alpha A| = \alpha^d |A|$  for  $d$ -dimensional matrices  $A$ , the term  $F(\alpha\theta)$  is

$$F(\alpha\theta) = F(\alpha v, \alpha M), \tag{23}$$

$$= \frac{d}{2} \log 2\pi - \frac{1}{2} \log \alpha^d | - 2M | - \frac{1}{4} (\alpha v)^T (\alpha M)^{-1} (\alpha v), \tag{24}$$

$$= \frac{d}{2} \log 2\pi - \frac{d}{2} \log \alpha - \frac{1}{2} \log | - 2M | - \frac{\alpha}{4} v^T M^{-1} v. \tag{25}$$

Similarly, we have

$$\alpha F(\theta) = \alpha F(v, M) = \frac{d\alpha}{2} \log 2\pi - \frac{\alpha}{2} \log | - 2M | - \frac{\alpha}{4} v^T M^{-1} v. \tag{26}$$

Thus, by subtracting equation (26) to equation (25), we obtain

$$F(\alpha\theta) - \alpha F(\theta) = \frac{d(1-\alpha)}{2} \log 2\pi - \frac{d}{2} \log \alpha - \frac{1-\alpha}{2} \log | - 2M |. \tag{27}$$

Therefore, we deduce that

$$F(\alpha\theta) - \alpha F(\theta) = \log \left( \frac{(2\pi)^{d\frac{1-\alpha}{2}}}{\alpha^{\frac{d}{2}}} | - 2M |^{-\frac{1-\alpha}{2}} \right) \tag{28}$$

$$= \log \left( \frac{(2\pi)^{\frac{d}{2}(1-\alpha)}}{\alpha^{\frac{d}{2}}} |\Sigma|^{\frac{1-\alpha}{2}} \right), \tag{29}$$

hence the result of equation (22). For 1D Gaussians with the standard deviation  $\sigma > 0$ , this yields

$$H_{\alpha,\beta}(\mu, \sigma) = \frac{1}{1-\beta} \left( \frac{(\sqrt{2\pi}\sigma)^{1-\beta}}{\alpha^{\frac{(1-\beta)}{2(1-\alpha)}}} - 1 \right). \tag{30}$$

Note that the differential Sharma–Mittal entropy of Gaussians may potentially be negative.

Figure 1 displays the plot of the Sharma–Mittal entropy for a  $4 \times 4$  covariance matrix set to  $4I$ , where  $I$  denotes the identity matrix.

We also report respectively the Rényi, Tsallis and Shannon entropies for multivariate Gaussians

$$R_\alpha(\mu, \Sigma) = \log \frac{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}{\alpha^{\frac{d}{2(1-\alpha)}}}, \tag{31}$$

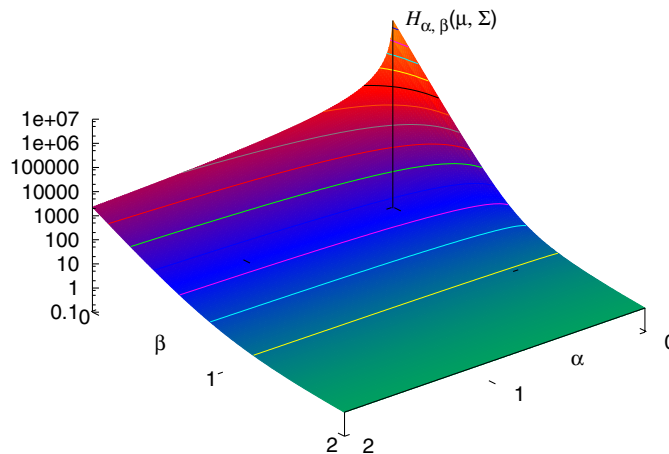
$$T_\alpha(\mu, \Sigma) = \frac{1}{1-\alpha} \left( \frac{((2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}})^{1-\alpha}}{\alpha^{\frac{d}{2}}} - 1 \right), \tag{32}$$

$$H(\mu, \Sigma) = \log \sqrt{(2\pi e)^d |\Sigma|} = \frac{1}{2} (d + d \log 2\pi + \log |\Sigma|). \tag{33}$$

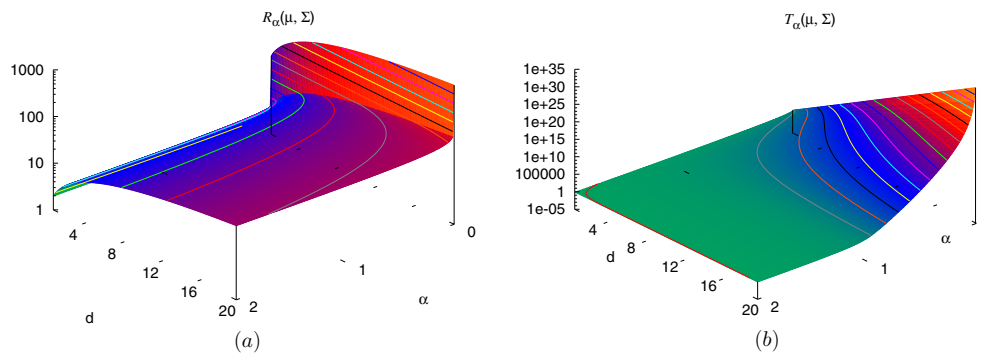
Figure 2 displays the plots of the Rényi (equation (31)) and Tsallis (equation (32)) entropies for a  $d \times d$ -dimensional covariance matrix  $\Sigma = \sigma^2 I = 4I$  ( $\sigma = 2$ ).

Information geometry [6] considers the underlying differential geometry induced by a divergence. From the Sharma–Mittal entropy, we can derive the Sharma–Mittal divergence [7] between two distributions  $P \sim p$  and  $Q \sim q$ :

$$D_{\alpha,\beta}(p : q) = \frac{1}{\beta - 1} \left( \left( \int p(x)^\alpha q(x)^{1-\alpha} dx \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \forall \alpha > 0, \alpha \neq 1, \beta \neq 1. \tag{34}$$



**Figure 1.** Plot of the Sharma–Mittal entropy (equation (22)) for the  $4 \times 4$  covariance matrix  $\Sigma = 4I$  (independent of the mean  $\mu$ ), where  $I$  denotes the identity matrix. Q3



**Figure 2.** Plot of the (a) Rényi (equation (31)) and (b) Tsallis (equation (32)) entropies for covariance matrices set to four times the matrix identity.

Note that  $D_{\alpha, \beta}(p : q) = 0$  if and only if  $p = q$ , since in that case  $\int p(x)^{\alpha} q(x)^{1-\alpha} dx = \int p(x) dx = 1$ .

For  $\alpha, \beta \rightarrow 1$ , the divergence tends to the renown Kullback–Leibler divergence. Let  $C_{\alpha}(p : q) = \int p(x)^{\alpha} q(x)^{1-\alpha} dx$  denote the  $\alpha$ -divergence [6], related to the Hellinger integral of order  $\alpha$ :  $H_{\alpha}(p, q) = 1 - C_{\alpha}(p, q)$ . For  $\alpha = \frac{1}{2}$ , the similarity measure  $C_{\frac{1}{2}}(p : q)$  is symmetric and called the Bhattacharyya coefficient. The Bhattacharyya coefficient is related to the following squared Hellinger distance:

$$H^2(p : q) = \frac{1}{2} \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx = 1 - C_{\frac{1}{2}}(p : q). \tag{35}$$

We rewrite compactly the Sharma–Mittal divergence of equation (34) as

$$D_{\alpha, \beta}(p : q) = \frac{1}{\beta - 1} (C_{\alpha}(p, q)^{\frac{1-\beta}{1-\alpha}} - 1). \tag{36}$$

□

Let us prove that for members  $p(x) = p_F(x|\theta)$  and  $q = p_F(x|\theta')$  belonging to the same exponential family  $\mathcal{E}_F$ , we have  $C_\alpha(p : q) = e^{-J_{F,\alpha}(\theta:\theta')}$ , where

$$J_{F,\alpha}(\theta : \theta') = \alpha F(\theta) + (1 - \alpha)F(\theta') - F(\alpha\theta + (1 - \alpha)\theta') \tag{37}$$

is a Jensen difference divergence [8].

**Proof.**

$$C_\alpha(p : q) = \int p_F(x|\theta)^\alpha p_F(x|\theta')^{1-\alpha} dx, \tag{38}$$

$$C_\alpha(\theta : \theta') = \int \exp^{\alpha(t(x,\theta)-F(\theta)+k(x))} \exp^{(1-\alpha)(t(x,\theta')-F(\theta')+k(x))} dx \tag{39}$$

$$= \int e^{t(x,\alpha\theta+(1-\alpha)\theta')} \exp^{-\alpha F(\theta)-(1-\alpha)F(\theta')+k(x)} dx \tag{40}$$

$$= \int e^{F(\alpha\theta+(1-\alpha)\theta')-\alpha F(\theta)-(1-\alpha)F(\theta')} p_F(x; \alpha\theta + (1 - \alpha)\theta') dx \tag{41}$$

$$= e^{-J_{F,\alpha}(\theta:\theta')} \int p_F(x; \alpha\theta + (1 - \alpha)\theta') dx \tag{42}$$

$$= e^{-J_{F,\alpha}(\theta:\theta')} > 0. \tag{43}$$

Observe that for  $\alpha \in (0, 1)$ ,  $\alpha\theta + (1 - \alpha)\theta' \in \Theta$  since  $\Theta$  is an open convex set, and therefore the distribution  $p_F(x; \alpha\theta + (1 - \alpha)\theta')$  is well defined in equation (41).

It follows that the Sharma–Mittal divergence of distributions belonging to the same exponential family (even when  $k(x) \neq 0$ ) is the following closed-form formula:

$$D_{\alpha,\beta}(p : q) = \frac{1}{\beta - 1} (C_\alpha(\theta_p : \theta_q)^{\frac{1-\beta}{1-\alpha}} - 1) \tag{44}$$

$$= \frac{1}{\beta - 1} (e^{-\frac{1-\beta}{1-\alpha}J_{F,\alpha}(\theta_p;\theta_q)} - 1). \tag{45}$$

For multivariate Gaussians, let us explicit the Jensen difference divergence  $J_{F,\alpha}$  as the difference of two terms using the  $(\mu, \Sigma)$  coordinate system

$$\alpha F(\theta) + (1 - \alpha)F(\theta') = \frac{d}{2} \log 2\pi + \frac{1}{2} \log |\Sigma|^\alpha |\Sigma'|^{1-\alpha} + \frac{\alpha}{2} \mu^T \Sigma^{-1} \mu + \frac{1 - \alpha}{2} \mu'^T \Sigma'^{-1} \mu' \tag{46}$$

and  $F(\alpha\theta + (1 - \alpha)\theta')$ . Let

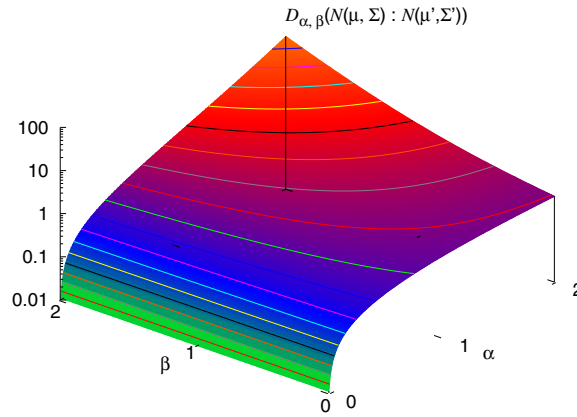
$$\bar{\theta}_\alpha = \alpha\theta + (1 - \alpha)\theta' = \left( \alpha \Sigma^{-1} \mu + (1 - \alpha) \Sigma'^{-1} \mu', -\frac{\alpha}{2} \Sigma^{-1} - \frac{1 - \alpha}{2} \Sigma'^{-1} \right) = (\bar{v}_\alpha, \bar{M}_\alpha). \tag{47}$$

Denote by  $\bar{\Sigma}_\alpha = -\frac{1}{2} \bar{M}_\alpha^{-1}$  and  $\bar{\mu}_\alpha = \bar{v}_\alpha \bar{\Sigma}_\alpha$  the corresponding parameters. Using equation (7), we have

$$F(\alpha\theta + (1 - \alpha)\theta') = F(\bar{\mu}_\alpha, \bar{\Sigma}_\alpha) = \frac{1}{2} \log(2\pi)^d |\bar{\Sigma}_\alpha| + \frac{1}{2} \bar{\mu}_\alpha^T \bar{\Sigma}_\alpha^{-1} \bar{\mu}_\alpha. \tag{48}$$

It follows that the Jensen difference divergence between two Gaussian distributions  $p \sim N(\mu, \Sigma)$  and  $q \sim N(\mu', \Sigma')$  is given by the closed-form formula

$$J_{F,\alpha}(p : q) = \frac{1}{2} \left( \log \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} + \alpha \mu^T \Sigma^{-1} \mu + (1 - \alpha) \mu'^T \Sigma'^{-1} \mu' - \bar{\mu}_\alpha^T \bar{\Sigma}_\alpha^{-1} \bar{\mu}_\alpha \right), \tag{49}$$



**Figure 3.** Plot of the Sharma–Mittal divergence  $D_{\alpha,\beta}$  (equation (54)) for univariate normal distributions with respective standard deviation  $\sigma = \sqrt{2}$  and  $\sigma' = 2$ , and mean difference  $\mu - \mu' = 4$ .

with

$$\bar{\Sigma}_\alpha = (\alpha \Sigma^{-1} + (1 - \alpha) \Sigma'^{-1})^{-1}, \tag{50}$$

$$\bar{\mu}_\alpha = \bar{\Sigma}_\alpha \bar{v}_\alpha = \bar{\Sigma} (\alpha \Sigma^{-1} \mu + (1 - \alpha) \Sigma'^{-1} \mu'). \tag{51}$$

Letting  $\Delta\mu = \mu' - \mu$ , equation (49) can further be rewritten compactly as

$$J_{F,\alpha}(p : q) = \frac{1}{2} \left( \log \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} + \alpha(1 - \alpha) \Delta\mu^T \bar{\Sigma}_\alpha^{-1} \Delta\mu \right). \tag{52}$$

Thus, we obtain a closed-form formula for the Sharma–Mittal divergence of multivariate Gaussians generalizing the Rényi  $\alpha$ -divergences, formerly reported in [9]:

$$D_{\alpha,\beta}(N(\mu, \Sigma) : N(\mu', \Sigma')) = \frac{1}{\beta - 1} \left( e^{-\frac{1-\beta}{2(1-\alpha)} \left( \log \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} + \alpha(1-\alpha) \Delta\mu^T \bar{\Sigma}_\alpha^{-1} \Delta\mu \right)} - 1 \right) \tag{53}$$

$$= \frac{1}{\beta - 1} \left( \left( \frac{|\Sigma|^\alpha |\Sigma'|^{1-\alpha}}{|\bar{\Sigma}_\alpha|} \right)^{-\frac{1-\beta}{2(1-\alpha)}} \exp \left( -\frac{\alpha(1-\beta)}{2} \Delta\mu^T \bar{\Sigma}_\alpha^{-1} \Delta\mu \right) - 1 \right). \tag{54}$$

Figure 3 shows the plot of the Sharma–Mittal divergence for univariate normal distributions with a respective standard deviation  $\sigma = \sqrt{2}$  and  $\sigma' = 2$ , and a mean difference  $\mu - \mu' = 4$  in equation (54). In practice, for numerical stability, we prefer to compute the divergence by first computing the Jensen difference divergence of equation (52) and then applying the generic formula of equation (45).

The underlying distribution is usually not explicitly given so that we first need to estimate the distribution or related quantities like its entropy [10]. Leonenko *et al* [11] proposed a method to estimate entropies using the  $k$ -nearest neighbor graph ( $k$ -NN) of an independently and identically distributed sample set  $x_1, \dots, x_n$ . However, their method suffers from the curse of dimensionality of computing  $k$ -NN graphs and falls short when dealing with moderate dimensions. For exponential families, we can estimate the natural parameter of an exponential

family using the maximum likelihood estimator that admits the unique global optimum [4, 5]  $\hat{\theta}$  such that

$$\nabla F(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n t(x_i). \quad (55)$$

For multivariate Gaussians, from the sufficient statistic  $t(x) = (x, x^T x)$ , we deduce that  $\nabla F(\hat{\theta}) = (\hat{\mu}, \hat{\Sigma} + \hat{\mu}\hat{\mu}^T) = (\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n x_i^T x_i)$ .

It follows a simple and fast scheme to estimate the Sharma–Mittal entropy (or divergence) from  $n$  observations sampled identically and independently from an exponential family distribution: estimate the natural parameter using equation (55) and apply the formula of equation (17).

To conclude, let us note that *any* arbitrary smooth density can be approximated by an exponential family of order depending on the approximation precision [12, 13] (enforcing no extra auxiliary carrier measure: that is, with  $k(x) = 0$ ). Thus, we can approximate the Sharma–Mittal entropy of an arbitrary probability density [13] by approximating it to a close exponential family, and then applying the closed-form formula (equation (17)). We believe that equations (17), (22), (45) (numerically stable) and (54) will prove to be useful when experimenting for suitable parameters  $(\alpha, \beta)$  in various statistical signal tasks [9].  $\square$

## Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions. FN (5793b870) would like to thank Dr Kitano and Dr Tokoro for their support. Additional material is available online at <http://www.informationgeometry.org/SharmaMittal/>

## References

- [1] Sharma B D and Mittal D P 1975 New nonadditive measures of inaccuracy *J. Math. Sci.* **10** 122–33
- [2] Aktürk E, Bağcı G B and Sever R 2007 Is Sharma–Mittal entropy really a step beyond Tsallis and Rényi entropies? arXiv:[cond-mat/0703277](https://arxiv.org/abs/cond-mat/0703277)
- [3] Üzengi Aktürk O, Aktürk E and Tomak M 2008 Can Sobolev inequality be written for Sharma–Mittal entropy? *Int. J. Theor. Phys.* **12** 3310–20
- [4] Brown L D 1986 *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory* (Hayworth, CA: Institute of Mathematical Statistics) ISBN:0-940-60010-2
- [5] Naudts J 2011 *Generalised Thermostatistics* (Berlin: Springer)
- [6] Amari S-I and Nagaoka H 2000 *Methods of Information Geometry* (Oxford: Oxford University Press)
- [7] Masi M 2005 A step beyond Tsallis and Rényi entropies *Phys. Lett. A* **338** 3–5
- [8] Burbea J and Rao C R 1982 On the convexity of some divergence measures based on entropy functions *IEEE Trans. Inf. Theory* **IT-28** 489–95
- [9] Hero A O, Ma B, Michel O and Gorman J D 2002 Applications of entropic spanning graphs *IEEE Signal Process. Mag.*
- [10] Poczós B and Schneider J 2011 On the estimation of  $\alpha$ -divergences *Proc. 14th Int. Conf. AI and Stat.*
- [11] Leonenko N, Pronzato L and Savani V 2008 A class of Rényi information estimators for multidimensional densities *Ann. Stat.* **36** 2153–82
- [12] Cobb L, Koppstein P and Chen N H 1983 Estimation and moment recursion relations for multimodal distributions of the exponential family *J. Am. Stat. Assoc.* **78** 124–30
- [13] Brody D C 2007 A note on exponential families of distributions *J. Phys. A: Math. Theor.* **78** 124–30

Q4

Q5

Q6

Q7



## QUERIES

### Page 1

#### Q1

Author: We have found some discrepancies in the source TeX file and reference PDF file. We have, however, followed the source TeX file as it seemed to be the updated one.

#### Q2

Author: Please be aware that the color figures in this article will only appear in color in the Web version. If you require color in the printed journal and have not previously arranged it, please contact the Production Editor now.

### Page 5

#### Q3

Author: Please provide better quality artwork for figures 1–3.

### Page 8

#### Q4

Author: Please check the details for any journal references that do not have a blue link as they may contain some incorrect information. Pale purple links are used for references to arXiv e-prints.

#### Q5

Author: Please check whether the book title and publisher's details (location and name) in reference [4] are okay as set.

#### Q6

Author: Please provide the volume number and page number in reference [9].

#### Q7

Author: Please provide the page range in reference [10].