

# Some results concerning maximum Rényi entropy distributions

Oliver Johnson<sup>a,\*</sup>, Christophe Vignat<sup>b</sup>

<sup>a</sup> *Statistical Laboratory, DPMMS, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK*  
<sup>b</sup> *L.I.S., 961 rue de la Houille Blanche, 38402 St. Martin d'Hères cedex, France*

Received 20 August 2005; received in revised form 10 May 2006; accepted 30 May 2006

Available online 19 January 2007

---

## Abstract

We consider the Student- $t$  and Student- $r$  distributions, which maximise Rényi entropy under a covariance condition. We show that they have information-theoretic properties which mirror those of the Gaussian distributions, which maximise Shannon entropy under the same condition. We introduce a convolution which preserves the Rényi maximising family, and show that the Rényi maximisers are the case of equality in a version of the Entropy Power Inequality. Further, we show that the Rényi maximisers satisfy a version of the heat equation, motivating the definition of a generalised Fisher information.

© 2006 Elsevier Masson SAS. All rights reserved.

## Résumé

Nous considérons les distributions de types Student- $t$  et Student- $r$  qui maximisent l'entropie de Rényi sous contrainte de covariance. Nous montrons qu'elles possèdent des propriétés informationnelles similaires à celles des distributions Gaussiennes, lesquelles maximisent l'entropie de Shannon sous la même contrainte. Nous montrons que ces distributions sont stables pour un certain type de convolution et qu'elles saturent une inégalité de la puissance entropique. De plus nous montrons que les lois à entropie de Rényi maximale vérifient une équation de la chaleur, ce qui permet de définir une information de Fisher généralisée.

© 2006 Elsevier Masson SAS. All rights reserved.

MSC: primary 94A17; secondary 60E99

Keywords: Entropy Power Inequality; Fisher information; Heat equation; Maximum entropy; Rényi entropy

---

## 1. Introduction

It is natural to ask whether the Shannon entropy of a  $n$ -dimensional random vector with density  $p$ , defined as

$$H(p) = - \int p(\mathbf{x}) \log p(\mathbf{x}) \, d\mathbf{x},$$

represents the only possible measure of uncertainty. For example, Rényi [17] introduces axioms on how we would expect such a measure to behave, and shows that these axioms are satisfied by a more general definition, as follows:

---

\* Corresponding author. Fax: +44 117 928 7999.

E-mail addresses: [O.Johnson@bristol.ac.uk](mailto:O.Johnson@bristol.ac.uk) (O. Johnson), [vignat@univ-mlv.fr](mailto:vignat@univ-mlv.fr) (C. Vignat).

**Definition 1.1.** Given a probability density  $p$  valued on  $\mathbb{R}^n$ , for  $q \neq 1$  define the  $q$ -Rényi entropy to be:

$$H_q(p) = \frac{1}{1-q} \log \left( \int p(\mathbf{x})^q \, d\mathbf{x} \right).$$

Note that by L'Hôpital's rule, since  $\frac{d}{dt} a^t = a^t \log_e a$ ,

$$\lim_{q \rightarrow 1} H_q(p) = \lim_{q \rightarrow 1} \frac{- \int p(\mathbf{x})^q \log p(\mathbf{x}) \, d\mathbf{x}}{\int p(\mathbf{x})^q \, d\mathbf{x}} = H(p). \quad (1)$$

As Gnedenko and Korolev [12] remark, under a variety of natural conditions the distributions which maximise Shannon entropy are well-known ones, with interesting properties. This paper gives parallels to some of these properties for the Rényi maximisers.

1. Under a covariance constraint Shannon entropy is maximised by the Gaussian distribution. In Proposition 1.3 we review the fact that under a covariance constraint Rényi entropy is maximised by Student distributions.
2. The Gaussians have the appealing property of stability (that is, given  $Z_1$  and  $Z_2$  independent Gaussians,  $Z_1 + Z_2$  is also Gaussian). In Definition 2.2, we introduce the  $\star$ -convolution, which generalises the addition operation. In Lemma 2.3, we extend the stability property by showing that if  $R_1$  and  $R_2$  are Rényi maximisers then so is  $R_1 \star R_2$ .
3. The Entropy Power Inequality (see Eq. (7) below) shows that the Gaussian represents the extreme case for how much entropy can change on addition. Theorem 2.4 gives the equivalent of an Entropy Power Inequality, with the Rényi maximisers playing an extremal role.
4. The Gaussian density satisfies the heat equation, which leads to a representation of Shannon entropy as an integral of Fisher Informations (known as the de Bruijn identity). In Theorem 3.1 we show that the Rényi densities satisfy a generalisation of the heat equation, and deduce what quantity must replace the Fisher information in general.

First, as in Costa, Hero and Vignat [5], we identify the Rényi maximising densities, which are Student- $t$  and Student- $r$  distributions, and review some of their properties which we will use later in the paper.

**Definition 1.2.** For  $n/(n+2) < q$  and  $q \neq 1$ , define the  $n$ -dimensional probability density  $g_{q,\mathbf{C}}$  as

$$g_{q,\mathbf{C}}(\mathbf{x}) = A_q (1 - (q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x})_+^{1/(q-1)} \quad (2)$$

with

$$\beta = \beta_q = \frac{1}{2q - n(1-q)},$$

and normalisation constants

$$A_q = \begin{cases} \left( \Gamma\left(\frac{1}{1-q}\right) (\beta(1-q))^{n/2} \right) / \left( \Gamma\left(\frac{1}{1-q} - \frac{n}{2}\right) \pi^{n/2} |\mathbf{C}|^{1/2} \right) & \text{if } \frac{n}{n+2} < q < 1, \\ \left( \Gamma\left(\frac{q}{q-1} + \frac{n}{2}\right) (\beta(q-1))^{n/2} \right) / \left( \Gamma\left(\frac{q}{q-1}\right) \pi^{n/2} |\mathbf{C}|^{1/2} \right) & \text{if } q > 1. \end{cases}$$

Here  $x_+ = \max(x, 0)$  denotes the positive part. We write  $\mathbf{R}_{q,\mathbf{C}}$  for a random variable with density  $g_{q,\mathbf{C}}$ , which has mean  $\mathbf{0}$  and covariance  $\mathbf{C}$ .

Notice that if we write  $\Omega_{q,\mathbf{C}}$  for the support of  $g_{q,\mathbf{C}}$ , then for  $q > 1$ ,  $\Omega_{q,\mathbf{C}} = \{\mathbf{x} : \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} \leq 2q/(q-1) + n\}$ , and for  $q < 1$ ,  $\Omega_{q,\mathbf{C}} = \mathbb{R}^n$ .

Note further that since

$$\lim_{q \rightarrow 1} \frac{\Gamma(1/(1-q))(1-q)^{n/2}}{\Gamma(1/(1-q) - n/2)} = 1 \quad \text{and} \quad \lim_{q \rightarrow 1} (1 - (q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x})_+^{1/(q-1)} = \exp(-\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}/2),$$

the limit

$$\lim_{q \rightarrow 1} g_{q,\mathbf{C}}(\mathbf{x}) = g_{1,\mathbf{C}}(\mathbf{x}) = ((2\pi)^n |\mathbf{C}|)^{-1/2} \exp(-\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}/2),$$

the Gaussian density. Throughout this paper, we write  $\mathbf{Z}_{\mathbf{C}}$  for a  $\mathcal{N}(\mathbf{0}, \mathbf{C})$  random variable.

We now state the maximum entropy property, as follows.

**Proposition 1.3.** *Given any  $q > n/(n + 2)$ , and positive definite symmetric matrix  $\mathbf{C}$ , among all probability densities  $f$  with mean  $\mathbf{0}$  and  $\int_{\Omega_{q,\mathbf{C}}} f(\mathbf{x})\mathbf{x}\mathbf{x}^T d\mathbf{x} = \mathbf{C}$ , the Rényi entropy is uniquely maximised by  $g_{q,\mathbf{C}}$ , that is*

$$H_q(f) \leq H_q(g_{q,\mathbf{C}}),$$

with equality if and only if  $f = g_{q,\mathbf{C}}$  almost everywhere.

**Proof.** See Section A.1.  $\square$

Throughout this paper, we write  $\chi_m$  for a random variable with density

$$f_m(x) = \frac{2^{1-m/2}}{\Gamma(m/2)} x^{m-1} \exp\left(-\frac{x^2}{2}\right), \quad \text{for } x > 0. \tag{3}$$

(Strictly speaking, this is only a  $\chi$  random variable when the parameter  $m$  is an integer, but it is simpler to adopt the convention of allowing non-integer  $m$  than to refer to the square root of a  $\Gamma(m/2)$  random variable with scale factor 2.)

We briefly review stochastic representations of the Rényi maximisers, which we will use throughout the paper. For the sake of completeness, we present proofs of these results in Section A.2. Part 1 of Proposition 1.4 follows for example from p. 393 of Eaton [10], Part 2 of Proposition 1.4 is stated in Dunnett [9], and Part 3 of this proposition is a multivariate version of a result stated as long ago as 1915 by Fisher [11].

**Proposition 1.4.** *Writing  $\mathbf{R}_{q,\mathbf{C}}$  for a  $n$ -dimensional  $q$ -Rényi maximiser with mean  $\mathbf{0}$  and covariance  $\mathbf{C}$ , and writing  $\mathbf{Z}_{\mathbf{C}}$  for a  $\mathcal{N}(\mathbf{0}, \mathbf{C})$ :*

(1) *Student- $r$ .* For any  $q > 1$ , writing  $m = n + 2q/(q - 1)$

$$\mathbf{R}_{q,\mathbf{C}}U \sim \mathbf{Z}_{m\mathbf{C}}, \tag{4}$$

where  $U \sim \chi_m$  (independent of  $\mathbf{R}_{q,\mathbf{C}}$ ).

(2) *Student- $t$ .* For any  $n/(n + 2) < q < 1$ , writing  $m = 2/(1 - q) - n > 2$ ,

$$\mathbf{R}_{q,\mathbf{C}} \sim \mathbf{Z}_{(m-2)\mathbf{C}}/U, \tag{5}$$

where  $U \sim \chi_m$  (independent of  $\mathbf{Z}$ ).

(3) *Duality.* Given matrix  $\mathbf{D}$ , define the map

$$\Theta_{\mathbf{D}}(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\mathbf{x}^T \mathbf{D}^{-1} \mathbf{x} + 1}}.$$

For  $q < 1$ , writing  $m = 2/(1 - q) - n$ , if  $\mathbf{R}_{q,\mathbf{C}}$  is a Rényi maximiser, then  $\Theta_{\mathbf{C}(m-2)}(\mathbf{R}_{q,\mathbf{C}}) \sim \mathbf{R}_{p,\mathbf{C}^*}$ , where

$$\frac{1}{p-1} = \frac{1}{1-q} - \frac{n}{2} - 1$$

(so  $q < 1$  implies that  $p > 1$ ) and  $\mathbf{C}^* = \mathbf{C}((m - 2)/(m + n))$ .

**Proof.** See Section A.2.  $\square$

Stochastic representations (4) and (5) can be used to compute the covariance and entropy of  $\mathbf{R}_{q,\mathbf{C}}$ . For example, for  $q < 1$ , since  $U \sim \chi_m$ , the  $\mathbb{E} \frac{1}{U^2} = \frac{1}{m-2}$ , so that

$$\text{Cov}(\mathbf{R}_{q,\mathbf{C}}) = \mathbb{E} \mathbf{Z}_{(m-2)\mathbf{C}} \mathbf{Z}_{(m-2)\mathbf{C}}^T \mathbb{E} \frac{1}{U^2} = (m - 2) \mathbf{C} \frac{1}{m - 2},$$

as claimed.

Similarly for  $q < 1$ , the Shannon entropy  $H_1(\mathbf{R}_{q,\mathbf{C}})$  is given by (writing  $m = 2/(1 - q) - n$ )

$$\begin{aligned} -\mathbb{E} \log g_{q,\mathbf{C}}(\mathbf{R}_{q,\mathbf{C}}) &= -\log A_q + \frac{m+n}{2} \mathbb{E} \log \left( 1 + \frac{\mathbf{Z}_{(m-2)\mathbf{C}}^T \mathbf{C}^{-1} \mathbf{Z}_{(m-2)\mathbf{C}}}{(m-2)U^2} \right) \\ &= -\log A_q + \frac{m+n}{2} \mathbb{E} \log \left( 1 + \frac{\mathbf{N}^T \mathbf{N}}{U^2} \right) \\ &= -\log A_q + \frac{m+n}{2} \mathbb{E} (\log \chi_{m+n}^2 - \log \chi_m^2), \end{aligned}$$

where  $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ , and since  $\mathbb{E} \log \chi_m^2 = \Psi(\frac{m}{2})$  where  $\Psi(\cdot)$  is the digamma function, we obtain

$$H_1(\mathbf{R}_{q,\mathbf{C}}) = -\log A_q + \frac{1}{1-q} \left( \Psi\left(\frac{1}{1-q}\right) - \Psi\left(\frac{1}{1-q} - \frac{n}{2}\right) \right). \tag{6}$$

**Remark 1.5.** Indeed, the theory of such stochastic representations can be generalised from the setting of [15] and [16] to multivariate maximisers with different powers. That is, given a positive sequence  $(p_1, \dots, p_n)$ , the solution to the problem

$$\max H_q(\mathbf{X}) \text{ such that } \mathbb{E}|X_i|^{p_i} = K_i$$

is a random vector  $\mathbf{X}$  with density given by

$$f(\mathbf{x}) \propto \left( 1 + \sum_{i=1}^n a_i |x_i|^{p_i} \right)_+^{1/(q-1)},$$

where it can be shown that the  $a_i$  all have the same sign as  $1 - q$ . Moreover, if  $\mathbf{X}$  is such a maximiser with  $q > 1$ , then for  $k = 1, \dots, n$  random variables  $Z_k = U_k^{1/p_k} X_k$  are independently power-exponential distributed with marginal densities

$$f(z_k) = \frac{p_k a_k^{1/p_k}}{2\Gamma(1/p_k)} \exp(a_k |z_k|^{p_k}), \quad a_k < 0,$$

when  $U_k$  is  $\chi$ -distributed with  $m = 2/(q - 1) + 2 + \sum_{i=1}^n 2/p_i$  degrees of freedom and independent of  $\mathbf{X}$ .

**2.  $\star$ -convolution and relative entropy**

In this section, we introduce a new operation, which we refer to as the  $\star$ -convolution. In Lemma 2.3 we show that this  $\star$ -convolution preserves the class of Rényi entropy maximisers, and in Theorem 2.4 show that it satisfies a version of the entropy power inequality.

We will say that a distribution is  $q$ -Rényi if it maximises the  $q$ -Rényi entropy. For the sake of simplicity, we write  $D(X\|Y) = D_1(f_X\|f_Y)$  for the relative entropy between the two densities  $f_X$  and  $f_Y$  of random variables  $X$  and  $Y$ . We define a new distance measure:

**Definition 2.1.** Given a  $n$ -dimensional random vector  $\mathbf{T}$  with mean  $\mathbf{0}$  and covariance  $\mathbf{C}$ , we define its distance from a  $n$ -dimensional  $q$ -Rényi maximiser  $\mathbf{R}_{q,\mathbf{C}}$  (for  $q > 1$ ) to be

$$d(\mathbf{T}|\mathbf{R}_{q,\mathbf{C}}) = D(\mathbf{T}U\|\mathbf{Z}),$$

where  $U$  is a  $\chi_m$  random variable (with  $m = n + 2q/(q - 1)$  degrees of freedom) independent of  $\mathbf{T}$ , and  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, m\mathbf{C})$ .

Note that  $d$  inherits positive definiteness from  $D$  – that is  $d(\mathbf{T}|\mathbf{R}_{q,\mathbf{C}}) \geq 0$ , with equality if and only if  $\mathbf{T} \sim \mathbf{R}_{q,\mathbf{C}}$ . Note further that Eq. (13) below implies that

$$d(\mathbf{T}|\mathbf{R}_{q,\mathbf{C}}) = D(\mathbf{T}U\|\mathbf{R}_{q,\mathbf{C}}U) \leq D(\mathbf{T}\|\mathbf{R}_{q,\mathbf{C}}).$$

Motivated by Proposition 1.4, we make the following definition:

**Definition 2.2.** For fixed  $q > 1$ , given two  $n$ -dimensional random vectors  $\mathbf{S}, \mathbf{T}$ , with covariance matrices  $\mathbf{C}_\mathbf{S}$  and  $\mathbf{C}_\mathbf{T}$ , define the  $\star_q$ -convolution (or just  $\star$ -convolution) of  $\mathbf{S}$  and  $\mathbf{T}$  to be the  $n$ -dimensional random vector

$$\begin{aligned} \mathbf{S} \star \mathbf{T} &= \Theta_{m\mathbf{C}} \left( \frac{U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T}}{V} \right) \\ &= \frac{(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T})}{\sqrt{(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T})^T(m\mathbf{C})^{-1}(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T}) + V^2}}, \end{aligned}$$

where  $\mathbf{C} = \mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T}$ , and  $U^{(S)}, U^{(T)}, V$  are independent  $\chi$  random variables, where  $U^{(S)}$  and  $U^{(T)}$  have  $m = n + 2q/(q - 1)$  degrees of freedom, and  $V$  has  $2q/(q - 1)$  degrees of freedom.

Again, notice that as  $q \rightarrow 1$ ,  $U^{(\cdot)}/(2q/(q - 1)) \rightarrow 1$  and  $V/(2q/(q - 1)) \rightarrow 1$  by the Law of Large Numbers, so  $\mathbf{S} \star \mathbf{T} \xrightarrow{d} \mathbf{S} + \mathbf{T}$ .

**Lemma 2.3.** For  $q > 1$ , if  $\mathbf{S}$  and  $\mathbf{T}$  are  $q$ -Rényi entropy maximisers with covariances  $\mathbf{C}_\mathbf{S}$  and  $\mathbf{C}_\mathbf{T}$  then  $\mathbf{S} \star \mathbf{T}$  is also a  $q$ -Rényi entropy maximiser, with covariance  $\mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T}$ .

**Proof.** By Proposition 1.4.1, writing  $m = n + 2q/(q - 1)$ , we know that  $U^{(S)}\mathbf{S}$  and  $U^{(T)}\mathbf{T}$  are  $\mathcal{N}(\mathbf{0}, m\mathbf{C}_\mathbf{S})$  and  $\mathcal{N}(\mathbf{0}, m\mathbf{C}_\mathbf{T})$  respectively. We define  $\tilde{q}$  by  $1/(1 - \tilde{q}) = 1 + 1/(q - 1) + n/2$ , and write  $\tilde{m} = 2/(1 - \tilde{q}) - n = 2q/(q - 1) = m - n$ . Then random variable  $\mathbf{W} = \sqrt{(\tilde{m} - 2)/m}(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T})$  is  $\mathcal{N}(\mathbf{0}, (\tilde{m} - 2)\mathbf{C})$ , where  $\mathbf{C} = \mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T}$ .

Then (by Proposition 1.4.2) since  $V$  has  $\tilde{m}$  degrees of freedom,  $\mathbf{W}/V$  is  $\tilde{q}$ -Rényi, with covariance  $\mathbf{C}$ . Finally (by Proposition 1.4.3),  $\Theta_{(\tilde{m}-2)\mathbf{C}}(\mathbf{W}/V)$  is  $q^*$ -Rényi, where  $1/(q^* - 1) = 1/(1 - q_1) - n/2 - 1 = 1/(q - 1)$ , so in fact it is  $q$ -Rényi with covariance  $\mathbf{C}(\tilde{m} - 2)/(\tilde{m} + n)$ . Hence,  $\mathbf{S} \star \mathbf{T} = \sqrt{m/(\tilde{m} - 2)} \Theta_{(\tilde{m}-2)\mathbf{C}}(\mathbf{W}/V)$  is  $q$ -Rényi with covariance  $\mathbf{C}m/(\tilde{m} + n) = \mathbf{C}$ , and the result follows.  $\square$

We now give a new ( $\star$ -convolution) version of the classical Entropy Power Inequality, which was first stated by Shannon as Theorem 15 of [18], with a ‘proof’ sketched in Appendix 6. More rigorous proofs appeared in Blachman [3] and later in Dembo, Cover and Thomas [8]. The result gives that for independent  $n$ -dimensional random vectors  $\mathbf{X}$  and  $\mathbf{Y}$ ,

$$\exp(2H(\mathbf{X} + \mathbf{Y})/n) \geq \exp(2H(\mathbf{X})/n) + \exp(2H(\mathbf{Y})/n), \tag{7}$$

with equality if and only if  $\mathbf{X}$  and  $\mathbf{Y}$  are Gaussian with proportional covariance matrices.

Writing  $\mathbf{C}_\mathbf{X}$  for the covariance matrix of  $\mathbf{X}$ , we know that  $D(\mathbf{X}|\mathbf{Z}_\mathbf{X}) = (n \log(2\pi e) + \log |\mathbf{C}_\mathbf{X}|)/2 - H(\mathbf{X})$ , so that the Entropy Power Inequality (7) is equivalent to

$$\begin{aligned} |\mathbf{C}_\mathbf{X} + \mathbf{C}_\mathbf{Y}|^{1/n} \exp(-2D(\mathbf{X} + \mathbf{Y}|\mathbf{Z}_{\mathbf{C}_\mathbf{X} + \mathbf{C}_\mathbf{Y}})/n) \\ \geq |\mathbf{C}_\mathbf{X}|^{1/n} \exp(-2D(\mathbf{X}|\mathbf{Z}_{\mathbf{C}_\mathbf{X}})/n) + |\mathbf{C}_\mathbf{Y}|^{1/n} \exp(-2D(\mathbf{Y}|\mathbf{Z}_{\mathbf{C}_\mathbf{Y}})/n). \end{aligned} \tag{8}$$

We give an equivalent of Eq. (8), with the  $\star$ -convolution replacing the operation of addition.

**Theorem 2.4.** Given  $q > 1$ , for independent  $n$ -dimensional random vectors  $\mathbf{S}, \mathbf{T}$  with mean  $\mathbf{0}$  and covariances  $\mathbf{C}_\mathbf{S}, \mathbf{C}_\mathbf{T}$ ,

$$\begin{aligned} |\mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T}|^{1/n} \exp(-2d(\mathbf{S} \star \mathbf{T}|\mathbf{R}_{q, \mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T}})/n) \\ \geq |\mathbf{C}_\mathbf{S}|^{1/n} \exp(-2d(\mathbf{S}|\mathbf{R}_{q, \mathbf{C}_\mathbf{S}})/n) + |\mathbf{C}_\mathbf{T}|^{1/n} \exp(-2d(\mathbf{T}|\mathbf{R}_{q, \mathbf{C}_\mathbf{T}})/n), \end{aligned}$$

with equality if and only if  $\mathbf{S}$  and  $\mathbf{T}$  are  $q$ -Rényi with proportional covariance matrices.

**Proof.** By Proposition A.5 below we know that for  $U^{(S)}, U^{(T)}, V, W$  all independent and  $\chi$ -distributed, where  $U^{(S)}, U^{(T)}, W$  have  $m = n + 2q/(q - 1)$  degrees of freedom, and  $V$  has  $2q/(q - 1)$  degrees of freedom:

$$\begin{aligned} d(\mathbf{S} \star \mathbf{T}|\mathbf{R}_{q, \mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T}}) &= D((\mathbf{S} \star \mathbf{T})W|\mathbf{Z}_{m(\mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T})}) \\ &= D\left(\frac{U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T}}{\sqrt{(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T})^T\mathbf{C}^{-1}(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T}) + V^2}}W\middle|\mathbf{Z}_{m(\mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T})}\right) \\ &\leq D(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T}|\mathbf{Z}_{m(\mathbf{C}_\mathbf{S} + \mathbf{C}_\mathbf{T})}). \end{aligned} \tag{9}$$

We can combine Eqs. (8) and (9) to obtain that

$$\begin{aligned} & |m\mathbf{C}_S + m\mathbf{C}_T|^{1/n} \exp(-2d(\mathbf{S} \star \mathbf{T} | \mathbf{R}_{q, \mathbf{C}_S + \mathbf{C}_T})/n) \\ & \geq |m\mathbf{C}_S + m\mathbf{C}_T|^{1/n} \exp(-2D(U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T} | \mathbf{Z}_{m(\mathbf{C}_S + \mathbf{C}_T)})/n) \\ & \geq |m\mathbf{C}_S|^{1/n} \exp(-2D(U^{(S)}\mathbf{S} | \mathbf{Z}_{m\mathbf{C}_S})/n) + |m\mathbf{C}_T|^{1/n} \exp(-2D(U^{(T)}\mathbf{T} | \mathbf{Z}_{m\mathbf{C}_T})/n) \\ & = |m\mathbf{C}_S|^{1/n} \exp(-2d(\mathbf{S} | \mathbf{R}_{q, \mathbf{C}_S})/n) + |m\mathbf{C}_T|^{1/n} \exp(-2d(\mathbf{T} | \mathbf{R}_{q, \mathbf{C}_T})/n), \end{aligned}$$

and the result follows. Equality holds in Eq. (9) if  $U^{(S)}\mathbf{S} + U^{(T)}\mathbf{T}$  is Gaussian. This, along with proportionality of covariance matrices, is also the condition for equality in Eq. (8).  $\square$

There is a parallel theory for the case  $q < 1$ , where we define a  $\circ$ -convolution:

**Definition 2.5.** For fixed  $q$  satisfying  $n/(n + 2) < q < 1$ , given two random vectors  $\mathbf{S}$  and  $\mathbf{T}$  with covariance matrices  $\mathbf{C}_S$  and  $\mathbf{C}_T$  respectively, define the  $\circ$ -convolution by

$$\mathbf{S} \circ \mathbf{T} = \Theta_{(m-2)(\mathbf{C}_S + \mathbf{C}_T)}^{-1} (\Theta_{(m-2)\mathbf{C}_S}(\mathbf{S}) \star_{\tilde{q}} \Theta_{(m-2)\mathbf{C}_T}(\mathbf{T}))$$

with  $m = 2/(1 - q) - n$ , where the  $\star$ -convolution is taken with respect to index  $\tilde{q}$  satisfying  $1/(\tilde{q} - 1) = m/2 - 1$  and

$$\Theta_{\mathbf{D}}^{-1}(\mathbf{X}) = \frac{\mathbf{X}}{\sqrt{1 - \mathbf{X}^T \mathbf{D}^{-1} \mathbf{X}}}.$$

This definition satisfies an analogue of Lemma 2.3:

**Lemma 2.6.** For  $q < 1$ , if  $\mathbf{S}$  and  $\mathbf{T}$  are  $q$ -Rényi entropy maximisers with covariances  $\mathbf{C}_S$  and  $\mathbf{C}_T$  then  $\mathbf{S} \circ \mathbf{T}$  is also a  $q$ -Rényi entropy maximiser, with covariance  $\mathbf{C}_S + \mathbf{C}_T$ .

**Proof.** By Proposition 1.4.3,  $\tilde{\mathbf{S}} = \Theta_{(m-2)\mathbf{C}_S}(\mathbf{S})$  maximises  $\tilde{q}$ -Rényi entropy with  $\tilde{q} > 1$  such that  $1/(\tilde{q} - 1) = 1/(1 - q) - n/2 - 1$ .

Moreover, the covariance matrix of  $\tilde{\mathbf{S}}$  is  $\mathbf{C}_{\tilde{S}} = \frac{m-2}{m+n} \mathbf{C}_S$ . The same result holds for  $\mathbf{T}$  and  $\mathbf{C}_{\tilde{T}} = \frac{m-2}{m+n} \mathbf{C}_T$ . As a consequence of Lemma 2.3,  $\tilde{\mathbf{S}} \star_{\tilde{q}} \tilde{\mathbf{T}}$  is a  $\tilde{q}$ -Rényi distribution with covariance  $\tilde{\mathbf{C}} = \mathbf{C}_{\tilde{S}} + \mathbf{C}_{\tilde{T}}$ .

Since by Proposition 1.4.3,  $\Theta_{(m-2)\mathbf{C}}(\mathbf{R}_{q, \mathbf{C}}) = \mathbf{R}_{\tilde{q}, \tilde{\mathbf{C}}}$ , where  $\tilde{\mathbf{C}} = (m - 2)\mathbf{C}/(m + n)$ , taking inverse maps,  $\Theta_{(m-2)\mathbf{C}}^{-1}(\mathbf{R}_{\tilde{q}, \tilde{\mathbf{C}}}) = \mathbf{R}_{q, \mathbf{C}}$ . Here  $\mathbf{C} = \tilde{\mathbf{C}}(m + n)/(m - 2) = (\mathbf{C}_{\tilde{S}} + \mathbf{C}_{\tilde{T}})(m + n)/(m - 2) = \mathbf{C}_S + \mathbf{C}_T$ , as required.  $\square$

### 3. $q$ -heat equation and $q$ -Fisher information

In this section, we show that the Rényi maximising distributions satisfy a version of the de Bruijn identity. That is, we can define a Fisher information quantity, and show in Eq. (11) that it is the derivative of entropy. First, we compute the exact constants in a result of Compte and Jou [4].

**Theorem 3.1.** For a fixed  $\mu$ , write  $f_\tau$  for the density of a  $\mathbf{R}_{q, \tau^\mu \mathbf{C}}$  random variable. If  $\mu = 2/(2 + n(q - 1)/2)$  then  $f_\tau$  satisfies a heat equation of the form

$$K_q \frac{\partial}{\partial \tau} f_\tau(\mathbf{x}) = \sum_{k,l} \mathbf{C}_{kl} \frac{\partial^2}{\partial x_k \partial x_l} f_\tau^q(\mathbf{x})$$

with

$$K_q = A_q^{q-1} \frac{2q(2 + n(q - 1))}{2q + n(q - 1)}.$$

**Proof.** By Eq. (2), we know that for a general choice of  $\mu$ :

$$f_\tau(\mathbf{x}) = \frac{A_q}{\tau^{n\mu/2}} \left( 1 - \frac{(q - 1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{1/(q-1)},$$

where

$$\beta = \frac{1}{2q - n(1 - q)}.$$

First note that

$$\frac{\partial}{\partial \tau} f_\tau(\mathbf{x}) = f_\tau(\mathbf{x}) \left( -\frac{n\mu}{2\tau} + \frac{\beta \mu \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^{\mu+1}} \left( 1 - \frac{(q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{-1} \right). \tag{10}$$

Further, for any  $k$ , writing  $\mathbf{A} = \mathbf{C}^{-1}$ :

$$\frac{\partial}{\partial x_k} f_\tau^q(\mathbf{x}) = \frac{A_q^q}{\tau^{nq\mu/2}} \left( 1 - \frac{(q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{1/(q-1)} \left( \frac{-2q\beta(\mathbf{A}\mathbf{x})_k}{\tau^\mu} \right).$$

Hence, for any  $k, l$ :

$$\begin{aligned} \frac{\partial^2}{\partial x_k \partial x_l} f_\tau^q(\mathbf{x}) &= \frac{A_q^q}{\tau^{nq\mu/2}} \left( 1 - \frac{(q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{1/(q-1)} \left( -\frac{2q\beta \mathbf{A}_{kl}}{\tau^\mu} \right) \\ &\quad + \frac{A_q^q}{\tau^{nq\mu/2}} \left( 1 - \frac{(q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{1/(q-1)-1} \left( \frac{4\beta^2 q}{\tau^{2\mu}} (\mathbf{A}\mathbf{x})_k (\mathbf{A}\mathbf{x})_l \right) \\ &= \frac{A_q^{q-1}}{\tau^{n(q-1)\mu/2}} f_\tau(\mathbf{x}) \left( \frac{-2q\beta \mathbf{A}_{kl}}{\tau^\mu} + \left( \frac{4q\beta^2 (\mathbf{A}\mathbf{x})_k (\mathbf{A}\mathbf{x})_l}{\tau^{2\mu}} \right) \left( 1 - \frac{(q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{-1} \right). \end{aligned}$$

Overall, we deduce that

$$\sum_{k,l} \mathbf{C}_{kl} \frac{\partial^2}{\partial x_k \partial x_l} f_\tau^q(\mathbf{x}) = \frac{A_q^{q-1}}{\tau^{n(q-1)\mu/2}} f_\tau(\mathbf{x}) \left( -\frac{2q\beta n}{\tau^\mu} + \frac{4q\beta^2 \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^{2\mu}} \left( 1 - \frac{(q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{\tau^\mu} \right)^{-1} \right)$$

so that equating this with Eq. (10) we obtain:

$$K_q = \frac{A_q^{q-1}}{\tau^{n(q-1)\mu/2+\mu-1}} \frac{4q\beta}{\mu}.$$

Now, we want this to not be a function of  $\tau$ , so take  $\mu = 2/(2 + n(q - 1))$ , and substitute for  $\beta$  to obtain

$$K_q = A_q^{q-1} \frac{2q(2 + n(q - 1))}{2q + n(q - 1)},$$

as claimed.  $\square$

Note that the value of the exponent  $\mu$  coincides with the one given by Compte and Jou [4]. Further, as  $\lim_{q \rightarrow 1} A_q^{q-1} = 1$ , so that  $\lim_{q \rightarrow 1} K_q = 2$ , as we would expect from the de Bruijn identity given in Lemma 2.2 of Johnson and Suhov [13].

We now evaluate the derivative of the Rényi entropy, extending the de Bruijn identity:

$$\begin{aligned} \frac{\partial}{\partial \tau} H_q(f_\tau) &= \frac{1}{1-q} \frac{(q-1) \int f_\tau(\mathbf{x})^{q-1} \frac{\partial}{\partial \tau} f_\tau(\mathbf{x}) \, d\mathbf{x}}{\int f_\tau(\mathbf{x})^q \, d\mathbf{x}} \\ &= -\frac{K_q^{-1}}{\int f_\tau(\mathbf{x})^q \, d\mathbf{x}} \sum_{k,l} \mathbf{C}_{kl} \int f_\tau(\mathbf{x})^{q-1} \frac{\partial^2}{\partial x_k \partial x_l} f_\tau^q(\mathbf{x}) \, d\mathbf{x} \\ &= \frac{K_q^{-1}}{\int f_\tau(\mathbf{x})^q \, d\mathbf{x}} \sum_{k,l} \mathbf{C}_{kl} \int \frac{\partial}{\partial x_l} f_\tau(\mathbf{x})^{q-1} \frac{\partial}{\partial x_k} f_\tau^q(\mathbf{x}) \, d\mathbf{x} \\ &= \frac{K_q^{-1} q (q-1)}{\int f_\tau(\mathbf{x})^q \, d\mathbf{x}} \sum_{k,l} \mathbf{C}_{kl} \int f_\tau(\mathbf{x})^{2q-3} \frac{\partial}{\partial x_l} f_\tau(\mathbf{x}) \frac{\partial}{\partial x_k} f_\tau(\mathbf{x}) \, d\mathbf{x} \\ &= K_q^{-1} q (q-1) \text{tr}(\mathbf{C}\mathbf{J}_q(f_\tau)), \end{aligned} \tag{11}$$

where we make the following definitions:

**Definition 3.2.** Given probability density  $p$ , define the  $q$ -score function

$$\rho_q(\mathbf{x}) = \nabla p(\mathbf{x})/p(\mathbf{x})^{2-q},$$

and the  $q$ -Fisher information matrix to be

$$\mathbf{J}_q(p) = \frac{\int p(\mathbf{x})\rho_q(\mathbf{x})\rho_q^T(\mathbf{x}) \, d\mathbf{x}}{\int p(\mathbf{x})^q \, d\mathbf{x}}.$$

Note that the numerator is the case  $p = 2, \lambda = q$  of the  $(p, \lambda)$  Fisher information introduced in Eq. (7) of [16]. We establish a multi-dimensional Cramér-Rao inequality:

**Proposition 3.3.** For the Fisher information  $\mathbf{J}_q$  defined above, given a random variable with density  $p$  and covariance  $\mathbf{C}$  then

$$\mathbf{J}_q(p) - \frac{\int p(\mathbf{x})^q \, d\mathbf{x}}{q^2} \mathbf{C}^{-1}$$

is positive definite, with equality if and only if  $p = g_{q,\mathbf{C}}$  everywhere.

**Proof.** The key is a Stein-like identity, as usual found using integration by parts, since writing  $\mathbf{A} = \mathbf{C}^{-1}$

$$\begin{aligned} \int p(\mathbf{x})(\rho_q(\mathbf{x}))_l(\mathbf{A}\mathbf{x})_k \, d\mathbf{x} &= \int \frac{\partial}{\partial x_l} p(\mathbf{x})p^{q-1}(\mathbf{x})(\mathbf{A}\mathbf{x})_k \, d\mathbf{x} \\ &= \frac{1}{q} \int \frac{\partial}{\partial x_l} (p^q(\mathbf{x})(\mathbf{A}\mathbf{x})_k) \, d\mathbf{x} \\ &= -\frac{1}{q} \int p^q(\mathbf{x})A_{kl} \, d\mathbf{x}. \end{aligned}$$

This means that for any real  $c$ , the positive definite matrix

$$\int p(\mathbf{x})(\rho_q(\mathbf{x}) + c\mathbf{A}\mathbf{x})(\rho_q(\mathbf{x}) + c\mathbf{A}\mathbf{x})^T \, d\mathbf{x} = \int p(\mathbf{x})\rho_q(\mathbf{x})\rho_q^T(\mathbf{x}) \, d\mathbf{x} + 2\frac{c}{q}\mathbf{A} \int p^q(\mathbf{x}) \, d\mathbf{x} + c^2\mathbf{A}.$$

So we choose  $c = (\int p^q(\mathbf{x}) \, d\mathbf{x})/q$ , and the result follows. Note that equality holds if and only if  $p = g_{q,\mathbf{C}}$  everywhere, since the Rényi maximiser has score function  $\rho(\mathbf{x}) = A_q^{q-1}(-2\beta)\mathbf{A}\mathbf{x}$ , and

$$\int g_{q,\mathbf{C}}^q(\mathbf{x}) \, d\mathbf{x}/q = \int g_{q,\mathbf{C}} A_q^{q-1}(1 - \beta(q - 1)\mathbf{x}^T\mathbf{A}\mathbf{x})/q \, d\mathbf{x} = A_q^{q-1}(1 - \beta(q - 1)n)/q = A_q^{q-1}(2\beta). \quad \square$$

Now, we can give the extensivity property for Fisher information defined in this way:

**Lemma 3.4.** For a compound system of independent random vectors  $\mathbf{X}$  and  $\mathbf{Y}$ , for  $q > 1/2$  the  $q$ -Fisher information satisfies:

$$\mathbf{J}_q(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} \alpha_q(\mathbf{Y})\mathbf{J}_q(\mathbf{X}) & 0 \\ 0 & \alpha_q(\mathbf{X})\mathbf{J}_q(\mathbf{Y}) \end{pmatrix},$$

where constant  $\alpha_q(\mathbf{X}) = (\int p_{\mathbf{X}}^{2q-1}(\mathbf{x}) \, d\mathbf{x})/(\int p_{\mathbf{X}}^q(\mathbf{x}) \, d\mathbf{x})$  and  $\alpha_q(\mathbf{Y})$  similarly.

**Proof.** We write  $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})$ , so that (omitting the arguments for clarity), we can express

$$\nabla p_{\mathbf{X},\mathbf{Y}} = (p_{\mathbf{Y}}\nabla p_{\mathbf{X}}, p_{\mathbf{X}}\nabla p_{\mathbf{Y}}).$$

Then

$$\begin{aligned} \iint p_{\mathbf{X},\mathbf{Y}}^{2q-3} \nabla p_{\mathbf{X},\mathbf{Y}} \nabla^T p_{\mathbf{X},\mathbf{Y}} &= \begin{pmatrix} \iint p_{\mathbf{Y}}^{2q-1} p_{\mathbf{X}}^{2q-3} \nabla p_{\mathbf{X}} \nabla^T p_{\mathbf{X}} & 0 \\ 0 & \iint p_{\mathbf{X}}^{2q-1} p_{\mathbf{Y}}^{2q-3} \nabla p_{\mathbf{Y}} \nabla^T p_{\mathbf{Y}} \end{pmatrix} \\ &= \begin{pmatrix} \int p_{\mathbf{Y}}^{2q-1} \int p_{\mathbf{X}}^q \mathbf{J}_q(\mathbf{X}) & 0 \\ 0 & \int p_{\mathbf{X}}^{2q-1} \int p_{\mathbf{Y}}^q \mathbf{J}_q(\mathbf{Y}) \end{pmatrix}, \end{aligned}$$



since for  $q > 1/2$ , the off-diagonal term

$$\left( \int p_{\mathbf{X}}^{2q-2} \nabla p_{\mathbf{X}} \right) \left( \int p_{\mathbf{Y}}^{2q-2} \nabla p_{\mathbf{Y}} \right) = \frac{1}{(2q-1)^2} \left( \int \nabla p_{\mathbf{X}}^{2q-1} \right) \left( \int \nabla p_{\mathbf{Y}}^{2q-1} \right) = 0,$$

since this is a perfect derivative, and since  $p_{\mathbf{X}}(\mathbf{x}) \rightarrow 0$  as  $\mathbf{x} \rightarrow \infty$ . The result follows since

$$\iint p_{\mathbf{X},\mathbf{Y}}^q = \left( \int p_{\mathbf{X}}^q \right) \left( \int p_{\mathbf{Y}}^q \right). \quad \square$$

**Acknowledgement**

This work was done during a visit by CV to OTJ at the Statistical Laboratory, University of Cambridge, in April 2005. CV would like to thank OTJ and the Statistical Laboratory for their hospitality.

**Appendix A. Proofs**

*A.1. Maximum entropy property*

In this section we give a proof of Proposition 1.3, which shows that  $g_{q,\mathbf{C}}$  are the Rényi entropy maximisers. The proof uses Lemma 1 of Lutwak, Yang and Zhang [16], which extends the classical Gibbs inequality, and is equivalent to Lemma A.2 below.

**Definition A.1.** For  $q \neq 1$ , given  $n$ -dimensional probability densities  $f$  and  $g$ , define the relative  $q$ -Rényi entropy distance from  $f$  to  $g$  to be

$$D_q(f\|g) = \frac{1}{1-q} \log \left( \int g^{q-1}(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} \right) + \frac{1-q}{q} H_q(g) - \frac{1}{q} H_q(f).$$

For  $q = 1$ , we write  $D_1(f\|g) = \int f(\mathbf{x}) \log(f(\mathbf{x})/g(\mathbf{x})) \, d\mathbf{x}$  for the standard relative entropy. We justify this as an extension by continuity; as  $q \rightarrow 1$ , as in (1),  $D_q(f\|g) \rightarrow -\int f(\mathbf{x}) \log g(\mathbf{x}) \, d\mathbf{x} - H_1(f) = D_1(f\|g)$ .

**Lemma A.2.** For any  $q > 0$ , and for any probability densities  $f$  and  $g$ , the relative entropy  $D_q(f\|g) \geq 0$ , with equality if and only if  $f = g$  almost everywhere.

**Proof.** The case  $q = 1$  is well-known. For  $q \neq 1$ , as in Lutwak, Yang and Zhang [16], the result is a direct application of Hölder’s inequality to  $\exp D_q(f\|g)$ . Although [16] only strictly speaking considers the 1-dimensional case, the general case is precisely the same.  $\square$

As with the Shannon maximisers, we use this Gibbs inequality Lemma A.2 to show that the densities of Definition 1.2 really do maximise the Rényi entropy.

**Proof of Proposition 1.3.** Since  $f$  and  $g_{q,\mathbf{C}}$  have the same covariance matrix,

$$\int_{\Omega_{q,\mathbf{C}}} (\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} = \int_{\Omega_{q,\mathbf{C}}} (\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}) g_{q,\mathbf{C}}(\mathbf{x}) \, d\mathbf{x}.$$

This means that for  $q \neq 1$

$$\begin{aligned} \int_{\Omega_{q,\mathbf{C}}} g_{q,\mathbf{C}}^{q-1}(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} &= \int_{\Omega_{q,\mathbf{C}}} A_q^{q-1} (1 - (q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\Omega_{q,\mathbf{C}}} A_q^{q-1} (1 - (q-1)\beta \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}) g_{q,\mathbf{C}}(\mathbf{x}) \, d\mathbf{x} \end{aligned}$$

$$= \int_{\Omega_{q,C}} g_{q,C}^q(\mathbf{x}) \, d\mathbf{x}. \tag{12}$$

For  $q = 1$ , the equivalent of the orthogonality property equation (12) is the well-known fact that

$$\int f(\mathbf{x}) \log g_{1,C}(\mathbf{x}) \, d\mathbf{x} = \int g_{1,C}(\mathbf{x}) \log g_{1,C}(\mathbf{x}) \, d\mathbf{x}.$$

Using Eq. (12) we simply evaluate

$$\begin{aligned} D_q(f \| g_{q,C}) &= \frac{1}{1-q} \log \left( \int g_{q,C}^{q-1}(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} \right) + \frac{1-q}{q} H_q(g_{q,C}) - \frac{1}{q} H_q(f) \\ &= \frac{1}{q} (H_q(g_{q,C}) - H_q(f)), \end{aligned}$$

and so the result follows by Lemma A.2.  $\square$

Note that this is an alternative proof to that given by Costa, Hero and Vignat [5], who introduced a non-symmetric directed divergence measure

$$D_q(f \| g) = \text{sign}(q - 1) \int_{\Omega_{q,C}} \frac{f^q(\mathbf{x})}{q} + \frac{q - 1}{q} g^q(\mathbf{x}) - f(\mathbf{x}) g^{q-1}(\mathbf{x}) \, d\mathbf{x}.$$

The approach of [5] is similar to that used by Cover and Thomas [6, p. 234] in the Gaussian case. The general theory of directed divergence measures is discussed by Csiszar [7] and by Ali and Silvey [1].

The paper [15] gives more general results concerning the maximum entropy property, in a more geometric context.

### A.2. Stochastic representation

**Proof of Proposition 1.4.** 1. By Eq. (3), since we take  $\beta(q - 1) = 1/m$  in Eq. (2), the density of  $\mathbf{R}_{q,C}U$  can be expressed as

$$\begin{aligned} g(\mathbf{y}) &= \frac{2^{1-m/2} A_q}{\Gamma(m/2)} \int_0^\infty \frac{1}{x^n} \left( 1 - \frac{\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}}{m x^2} \right)^{1/(q-1)} x^{m-1} \exp\left(-\frac{x^2}{2}\right) \, dx \\ &= \frac{2^{1-m/2}}{\Gamma(m/2)} A_q \exp\left(-\frac{\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}}{2m}\right) K. \end{aligned}$$

Here since  $m - n - 2 = 2/(q - 1)$ , taking  $u^2 = x^2 - \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}/m$ , so  $u \, du = x \, dx$ :

$$\begin{aligned} K &= \int_0^\infty \left( 1 - \frac{\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}}{m x^2} \right)^{1/(q-1)} x^{m-n-2} \exp\left(-\frac{x^2}{2} + \frac{\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}}{2m}\right) x \, dx \\ &= \int_0^\infty u^{2/(q-1)} \exp\left(-\frac{u^2}{2}\right) u \, du = 2^{1/(q-1)} \Gamma\left(\frac{q}{q-1}\right), \end{aligned}$$

and the result follows, since the constant

$$\frac{2^{1-m/2}}{\Gamma(m/2)} A_q K = \frac{1}{(2\pi m)^{n/2} |\mathbf{C}|^{1/2}},$$

since  $1 - m/2 + 1/(q - 1) = -n/2$  and  $\beta(q - 1) = 1/m$ .

2. In the same way, the density of  $\mathbf{Z}_{(m-2)C}/U$  can be expressed as

$$\begin{aligned} & \frac{2^{1-m/2}}{\Gamma(m/2)} \int_0^\infty \frac{x^n}{\sqrt{(2\pi(m-2))^n |\mathbf{C}|}} \exp\left(-\frac{x^2 \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}}{2(m-2)}\right) x^{m-1} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \frac{2^{1-m/2-n/2}}{\Gamma(m/2)} \sqrt{\frac{\beta(1-q)}{\pi^n |\mathbf{C}|}} \int_0^\infty \exp\left(-\frac{(1+d)x^2}{2}\right) x^{n+m-1} dx \\ &= \frac{\Gamma((n+m)/2)}{\Gamma(m/2)} \sqrt{\frac{\beta(1-q)}{\pi^n |\mathbf{C}|}} (1+d)^{-(n+m)/2} \\ &= \frac{\Gamma(1/(1-q))}{\Gamma(1/(1-q) - n/2)} \sqrt{\frac{\beta(1-q)}{\pi^n |\mathbf{C}|}} \left(1 + \frac{\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}}{m-2}\right)^{1/(q-1)}, \end{aligned}$$

writing  $d = (\mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}) / (m - 2)$ , and using the facts that  $(m + n)/2 = 1/(1 - q)$  and  $1/(m - 2) = \beta(1 - q)$ , the result follows.

3. For this choice of parameters,  $\mathbf{X} = \mathbf{R}_{q,\mathbf{C}}$  has density  $A_q (1 + \mathbf{x}^T \mathbf{D}^{-1} \mathbf{x})^{1/(q-1)}$ .

If  $\mathbf{Y} = \Theta_{\mathbf{D}}(\mathbf{X})$ , we can calculate the Jacobian  $|\partial \mathbf{X}|/|\partial \mathbf{Y}| = (1 - \mathbf{Y}^T \mathbf{D}^{-1} \mathbf{Y})^{-1-n/2}$ . Then, the standard change-of-variables relation gives that, since  $1 - \mathbf{Y}^T \mathbf{D}^{-1} \mathbf{Y} = (1 + \mathbf{X}^T \mathbf{D}^{-1} \mathbf{X})^{-1}$ , we know that  $\mathbf{Y}$  has density

$$g_{\mathbf{Y}}(\mathbf{y}) = (1 - \mathbf{y}^T \mathbf{D}^{-1} \mathbf{y})^{-1-n/2} g_{\mathbf{X}}(\Theta_{\mathbf{D}}^{-1}(\mathbf{y})).$$

Thus, in particular, taking  $\mathbf{X} \sim \mathbf{R}_{q,\mathbf{C}}$  and  $\mathbf{D} = \mathbf{C}(m - 2)$ , we know that

$$\begin{aligned} g_{\mathbf{Y}}(\mathbf{y}) &= (1 - \mathbf{y}^T \mathbf{D}^{-1} \mathbf{y})^{-1-n/2} A_q (1 - \mathbf{y}^T \mathbf{D}^{-1} \mathbf{y})^{-1/(q-1)} \\ &= A_q (1 - \mathbf{y}^T \mathbf{D}^{-1} \mathbf{y})^{1/(p-1)}. \end{aligned}$$

Since  $p > 1$ , we know that  $\mathbf{Y}$  has covariance  $\mathbf{D} \beta_p (p - 1) = \mathbf{D}(2p/(p - 1) + n)^{-1} = \mathbf{C}(m - 2)/(m + n)$ .

Further

$$\begin{aligned} A_q &= (\beta_q (1 - q))^{n/2} \Gamma\left(\frac{1}{1 - q}\right) / \left(\Gamma\left(\frac{1}{1 - q} - \frac{n}{2}\right) \pi^{n/2} |\mathbf{C}|^{1/2}\right) \\ &= (\beta_p (p - 1))^{n/2} \Gamma\left(\frac{1}{p - 1} + \frac{n}{2} + 1\right) / \left(\Gamma\left(\frac{1}{p - 1} + 1\right) \pi^{n/2} |(m - 2)/(m + n) \mathbf{C}|^{1/2}\right) = A_p, \end{aligned}$$

as required.  $\square$

Note that an alternate, stochastic proof of Eq. (4) can be deduced from the polar factorisation property of Student- $r$  vectors (see [2] for a detailed study): if  $\mathbf{X}$  is orthogonally invariant and  $\mathbf{X} = r\mathbf{U}$  where  $\mathbf{U}$  is uniformly distributed on the sphere, then  $r = \|\mathbf{X}\|$  and  $\mathbf{U} = \mathbf{X}/\|\mathbf{X}\|$  are independent. Since  $\mathbf{R}_{q,\mathbf{C}}$  is the marginal of a vector  $\mathbf{U}$  uniformly distributed on the sphere, we deduce that

$$\mathbf{R}_{q,\mathbf{C}} = \frac{\sqrt{m} \mathbf{C}^{1/2} \mathbf{Z}}{\sqrt{\mathbf{Z}^T \mathbf{Z} + \chi_{m-n}^2}},$$

where  $\mathbf{Z}$  is a Gaussian vector, and where random variable  $\sqrt{\mathbf{Z}^T \mathbf{Z} + \chi_{m-n}^2}$  is chi distributed with  $m$  degrees of freedom and independent of  $\mathbf{R}_{q,\mathbf{C}}$ . Thus, multiplying  $\mathbf{R}_{q,\mathbf{C}}$  by an independent chi-distributed random variable with  $m$  degrees of freedom yields a Gaussian vector with covariance matrix  $m\mathbf{C}$ , which is exactly Eq. (4).

### A.3. Projection results

To prove the Entropy Power Inequality, Theorem 2.4, we prove a technical result, Proposition A.5. This relies on two well-known results, Lemmas A.3 and A.4. Firstly as a consequence of the chain rule for relative entropy (see for example Theorem 2.5.3 of Cover and Thomas [6]):

**Lemma A.3.** For pairs of random variables  $(X, Y)$  and  $(U, V)$ ,

$$D((X, Y) \parallel (U, V)) \geq D(X \parallel U).$$

Equality holds if and only if for each  $x$ , the random variables  $Y|X=x$  and  $V|U=x$  have the same distribution. In particular if  $(X, Y)$  and  $(U, V)$  are independent pairs, equality holds if and only if  $Y$  and  $V$  have the same distribution.

Secondly, we recall a projection identity, first stated as Corollary 4.1 of [14]:

**Lemma A.4.** For random vectors  $\mathbf{X}$  and  $\mathbf{Y}$ , and for any invertible function  $\Phi$ :

$$D(\Phi(\mathbf{X}) \parallel \Phi(\mathbf{Y})) = D(\mathbf{X} \parallel \mathbf{Y}).$$

**Proposition A.5.** For a  $n$ -dimensional random vector  $\mathbf{M}$ , take  $N \sim \chi_{2q/(q-1)}$  and  $U \sim \chi_{2q/(q-1)+n}$ , where  $(\mathbf{M}, N, U)$  are independent:

$$D\left(\frac{\mathbf{M}}{\sqrt{\mathbf{M}^T \mathbf{C}^{-1} \mathbf{M} + N^2}} U \parallel \mathbf{Z}_C\right) \leq D(\mathbf{M} \parallel \mathbf{Z}_C),$$

where equality holds if  $\mathbf{M}$  is  $\mathcal{N}(\mathbf{0}, \mathbf{C})$ .

**Proof.** By combining Lemmas A.3 and A.4, if random variables  $Q$  and  $S$  have the same distribution and  $(\mathbf{P}, Q)$  and  $(\mathbf{R}, S)$  each form independent pairs then

$$D(\mathbf{P}Q \parallel \mathbf{R}S) \leq D((\mathbf{P}Q, Q) \parallel (\mathbf{R}S, S)) = D((\mathbf{P}, Q) \parallel (\mathbf{R}, S)) = D(\mathbf{P} \parallel \mathbf{R}). \quad (13)$$

Now, we define  $Y \sim \chi_{2q/(q-1)}$  and  $V \sim \chi_{2q/(q-1)+n}$ , both independent of  $\mathbf{Z}_C$ , so that  $U$  and  $V$  have the same distribution, as do  $N$  and  $Y$ . The LHS of the proposition becomes:

$$D\left(\frac{\mathbf{M}}{\sqrt{\mathbf{M}^T \mathbf{C}^{-1} \mathbf{M} + N^2}} U \parallel \frac{\mathbf{Z}_C}{\sqrt{\mathbf{Z}_C^T \mathbf{C}^{-1} \mathbf{Z}_C + Y^2}} V\right) \leq D\left(\frac{\mathbf{M}}{\sqrt{\mathbf{M}^T \mathbf{C}^{-1} \mathbf{M} + N^2}} \parallel \frac{\mathbf{Z}_C}{\sqrt{\mathbf{Z}_C^T \mathbf{C}^{-1} \mathbf{Z}_C + Y^2}}\right) \quad (14)$$

$$= D(\Theta_C(\mathbf{M}/N) \parallel \Theta_C(\mathbf{Z}_C/Y))$$

$$= D(\mathbf{M}/N \parallel \mathbf{Z}_C/Y) \quad (15)$$

$$\leq D(\mathbf{M} \parallel \mathbf{Z}_C), \quad (16)$$

and the result follows. Here Eq. (14) follows by Eq. (13), Eq. (15) follows by Lemma A.4 and Eq. (16) again follows by Eq. (13).  $\square$

## References

- [1] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, J. Roy. Statist. Soc. Ser. B 28 (1966) 131–142.
- [2] F. Barthe, M. Csörnyei, A. Naor, A note on simultaneous polar and Cartesian decomposition, in: Geometric Aspects of Functional Analysis, in: Lecture Notes in Math., vol. 1807, Springer, Berlin, 2003, pp. 1–19.
- [3] N.M. Blachman, The convolution inequality for entropy powers, IEEE Trans. Inform. Theory 11 (1965) 267–271.
- [4] A. Compte, D. Jou, Non-equilibrium thermodynamics and anomalous diffusion, J. Phys. A 29 (15) (1996) 4321–4329.
- [5] J. Costa, A. Hero, C. Vignat, On solutions to multivariate maximum  $\alpha$ -entropy problems, in: A. Rangarajan, M. Figueiredo, J. Zerubia (Eds.), EMMCVPR 2003, Lisbon, 7–9 July 2003, in: Lecture Notes in Computer Science, vol. 2683, Springer-Verlag, Berlin, 2003, pp. 211–228.
- [6] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley, New York, 1991.
- [7] I. Csizsár, Information-type measures of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar. 2 (1967) 299–318.
- [8] A. Dembo, T.M. Cover, J.A. Thomas, Information theoretic inequalities, IEEE Trans. Inform. Theory 37 (6) (1991) 1501–1518.
- [9] C.W. Dunnett, M. Sobel, A bivariate generalization of Student's  $t$ -distribution, with tables for certain special cases, Biometrika 41 (1954) 153–169.
- [10] M.L. Eaton, On the projections of isotropic distributions, Ann. Statist. 9 (2) (1981) 391–400.
- [11] R.A. Fisher, Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika 10 (1915) 507–521.
- [12] B.V. Gnedenko, V.Y. Korolev, Random Summation: Limit Theorems and Applications, CRC Press, Boca Raton, FL, 1996.

- [13] O.T. Johnson, Y.M. Suhov, Entropy and random vectors, *J. Statist. Phys.* 104 (1) (2001) 147–167.
- [14] S. Kullback, R. Leibler, On information and sufficiency, *Ann. Math. Statist.* 22 (1951) 79–86.
- [15] E. Lutwak, D. Yang, G. Zhang, Moment-entropy inequalities, *Ann. Probab.* 32 (1B) (2004) 757–774.
- [16] E. Lutwak, D. Yang, G. Zhang, Cramer–Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information, *IEEE Trans. Inform. Theory* 51 (2005) 473–478.
- [17] A. Rényi, On measures of entropy and information, in: J. Neyman (Ed.), *Proceedings of the 4th Berkeley Conference on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1961, pp. 547–561.
- [18] C.E. Shannon, W.W. Weaver, *A Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1949.