Journal of Inequalities in Pure and Applied Mathematics

http://jipam.vu.edu.au/

Volume 5, Issue 4, Article 102, 2004

ESTIMATORS, ESCORT PROBABILITIES, AND ϕ-EXPONENTIAL FAMILIES IN STATISTICAL PHYSICS

JAN NAUDTS

DEPARTEMENT NATUURKUNDE
UNIVERSITEIT ANTWERPEN
UNIVERSITEITSPLEIN 1
2610 ANTWERPEN, BELGIUM.
Jan.Naudts@ua.ac.be

Received 23 February, 2004; accepted 11 November, 2004
Communicated by C.E.M. Pearce

ABSTRACT. The lower bound of Cramér and Rao is generalized to pairs of families of probability distributions, one of which is escort to the other. This bound is optimal for certain families, called ϕ-exponential in the paper. Their dual structure is explored. They satisfy a variational principle with respect to an appropriately chosen entropy functional, which is the dual of a free energy functional.

Key words and phrases: Escort probability, Lower bound of Cramér and Rao, Generalized exponential family, Statistical manifold, Nonextensive thermostatistics.

2000 *Mathematics Subject Classification.* 82B30, 62H12.

1. INTRODUCTION

The aim of this paper is to translate some new results of statistical physics into the language of statistics. It is well-known that the exponential family of probability distribution functions (pdfs) plays a central role in statistical physics. When Gibbs [6] introduced the canonical ensemble in 1901 he postulated a distribution of energies E of the form

$$p(E) = \exp(G - \beta E), \tag{1.1}$$

where G is a normalization constant and where the control parameter β is the inverse temperature. Only recently [17], a proposal was made to replace (1.1) by a more general family of pdfs. The resulting domain of research is known under the name of Tsallis' thermostatistics. Some of the pdfs of Tsallis' thermostatistics are known in statistics under the name of Amari's α-family [3]. The latter have been introduced in the context of geometry of statistical manifolds [8]. The appearance of the same family of pdfs in both domains is not accidental. The apparent link between both domains is clarified in the present paper.

ISSN (electronic): 1443-5756

I am thankful to S. Abe who urged me to study the geometry of statistical distributions. I thank Dr. Ch. Vignat for pointing out ref. [13].
038-04

The new notion introduced in Tsallis' thermostatistics is that of pairs of families of pdfs, one of which is the *escort* of the other [4]. Some basic concepts of statistics can be generalized by replacing at well-chosen places the pdf by its escort. In particular, we show in the next section how to generalize Fisher's information and, correspondingly, how to generalize the well-known lower bound of Cramér and Rao. Section 3 studies the statistical manifold of a family for which there exists an escort family satisfying the condition under which the generalized Cramér-Rao bound is optimal. This optimizing family has an affine geometry. Since this is usually the characteristic property of an exponential family a generalization of the latter seems indicated.

Section 4 shows how a strictly positive non-decreasing function ϕ of $\mathbb{R}_+$ determines a function which shares some properties with the natural logarithm and therefore is called below a ϕ-logarithm. The inverse function is called the ϕ-exponential. In Section 5 it is used to define the ϕ-exponential family in the obvious way, by replacing the exponential function $\exp$ by the ϕ-exponential function. The standard exponential family is then recovered by the choice $\phi(x) = x$, the α-family of Amari by $\phi(x) = x^{(1+\alpha)/2}$, the equilibrium pdfs of Tsallis' thermostatistics by the choice $\phi(x) = x^q$.

The next three sections are used to establish the dual parametrization of the ϕ-exponential family and to discover the role of entropy functionals. Section 6 introduces a divergence of the Bregman type. In Section 7 it is used to prove the existence of an information function (or entropy functional) which is maximized by the ϕ-exponential pdfs. Section 8 introduces dual parameters — in statistical physics these are energy and temperature. The paper ends with a short discussion in Section 9.

There have been already some attempts to study Tsallis' thermostatistics from a geometrical point of view. Trasarti-Battistoni [15] conjectured a deep connection between non-extensivity and geometry. He also gives general references to the use of geometric ideas in statistical physics. Several authors [1, 16, 14] have introduced a divergence belonging to Csiszár's class of f-divergences, which leads to a generalization of the Fisher information metric adapted to the context of Tsallis' thermostatistics. The relation with the present work is unclear since here the geometry is determined by a divergence of the Bregman type. Also the recent work of Abe [2] seems to be unrelated.

2. ESTIMATORS AND ESCORT PDFS

Fix a measure space Ω, μ. Let $\mathcal{M}_1(\mu)$ denote the convex set of all probability distribution functions (pdfs) p normalized w.r.t. μ

$$\int_\Omega \mathrm{d}\mu(x)\, p(x) = 1. \tag{2.1}$$

Expectations w.r.t. p are denoted by $\mathbb{E}_p$

$$\mathbb{E}_p f = \int_\Omega \mathrm{d}\mu(x)\, p(x) f(x).$$

Fix an open domain D of $\mathbb{R}^n$. Consider a family of pdfs p_θ, parametrized with θ in D. The notation $\mathbb{E}_\theta$ will be used instead of $\mathbb{E}_{p_\theta}$. Simultaneously, a second family of pdfs $(P_\theta)_{\theta\in D}$ is considered. It is called the *escort family*. The notation $\mathbb{F}_\theta$ will be used instead of $\mathbb{E}_{P_\theta}$.

Recall that the Fisher information is given by

$$\begin{aligned} I_{kl}(\theta) &= \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta^k}\log(p_\theta)\right)\left(\frac{\partial}{\partial\theta^l}\log(p_\theta)\right) \\ &= \int_\Omega \mathrm{d}\mu(x)\, \frac{1}{p_\theta(x)}\frac{\partial p_\theta}{\partial\theta^k}\frac{\partial p_\theta}{\partial\theta^l}. \end{aligned} \tag{2.2}$$

A generalization, involving the two families of pdfs, is

$$g_{kl}(\theta) = \int_\Omega \mathrm{d}\mu(x)\, \frac{1}{P_\theta(x)} \frac{\partial p_\theta}{\partial \theta^k} \frac{\partial p_\theta}{\partial \theta^l}. \tag{2.3}$$

Clearly, the expression coincides with (2.2) if $P_\theta = p_\theta$.

The following definition is a slight generalization of the usual definition of an unbiased estimator.

Definition 2.1. An *estimator* of the family $(p_\theta)_{\theta\in D}$ is a vector of random variables c_k with the property that there exists a function F such that

$$\mathbb{E}_\theta c_k = \frac{\partial}{\partial \theta^k} F(\theta), \qquad k = 1, \ldots, n.$$

The function F will be called the *scale function* of the estimator.

The estimator is unbiased if $F(\theta) = \frac{1}{2}\theta_k\theta^k$ so that $\mathbb{E}_\theta c_k = \theta_k$. The well-known lower bound of Cramér and Rao can be written as

$$\frac{u^k u^l \left[\mathbb{E}_\theta c_k c_l - \left(\mathbb{E}_\theta c_k\right)\left(\mathbb{E}_\theta c_l\right)\right]}{\left[u^k v^l \frac{\partial^2 F}{\partial\theta^k \partial\theta^l}\right]^2} \geq \frac{1}{v^k v^l I_{kl}(\theta)},$$

for arbitrary u and v in $\mathbb{R}^n$.

A similar lower bound, involving the information matrix g_{kl} instead of Fisher's I_{kl}, is now formulated.

Theorem 2.1. *Let be given two families of pdfs $(p_\theta)_{\theta\in D}$ and $(P_\theta)_{\theta\in D}$ and corresponding expectations $\mathbb{E}_\theta$ and $\mathbb{F}_\theta$. Let c be an estimator of $(p_\theta)_{\theta\in D}$, with scale function F. Assume that the regularity condition*

$$\mathbb{F}_\theta \frac{1}{P_\theta(x)} \frac{\partial}{\partial\theta^k} p_\theta(x) = 0 \tag{2.4}$$

holds. Let $g_{kl}(\theta)$ be the information matrix introduced before. Then, for all u and v in $\mathbb{R}^n$ is

$$\frac{u^k u^l \left[\mathbb{F}_\theta c_k c_l - \left(\mathbb{F}_\theta c_k\right)\left(\mathbb{F}_\theta c_l\right)\right]}{\left[u^k v^l \frac{\partial^2}{\partial\theta^l\partial\theta^k} F(\theta)\right]^2} \geq \frac{1}{v^k v^l g_{kl}(\theta)}. \tag{2.5}$$

The bound is optimal (in the sense that equality holds whenever $u = v$) if there exist a normalization function $Z > 0$ and a function G such that

$$\frac{\partial}{\partial\theta^k} p_\theta(x) = Z(\theta) P_\theta(x) \frac{\partial}{\partial\theta^k}\left[G(\theta) - \theta^l c_l(x)\right] \tag{2.6}$$

holds for all k in $[1, \ldots, m]$, for all $\theta \in D$, and for μ-almost all x. In that case, c is an estimator of $(P_\theta)_{\theta\in D}$ with scale function G

$$\mathbb{F}_\theta c_k = \frac{\partial G}{\partial\theta^k}.$$

Proof. Let

$$X_k = \frac{1}{P_\theta}\frac{\partial}{\partial\theta^k} p_\theta \qquad \text{and} \qquad Y_k = c_k - \mathbb{F}_\theta c_k.$$

From Schwartz's inequality follows

$$\left(\mathbb{F}_\theta u^k Y_k v^l X_l\right)^2 \leq \left(\mathbb{F}_\theta u^k Y_k u^l Y_l\right)\left(\mathbb{F}_\theta v^k X_k v^l X_l\right).$$

The l.h.s. equals, using (2.4),

$$\begin{aligned}\left(\mathbb{F}_\theta u^k Y_k v^l X_l\right)^2 &= \left(u^k v^l \frac{\partial}{\partial\theta^l}\mathbb{E}_\theta c_k\right)^2 \\ &= \left(u^k v^l \frac{\partial^2}{\partial\theta^l\partial\theta^k}F(\theta)\right)^2.\end{aligned}$$

The first factor of the r.h.s. equals

$$\mathbb{F}_\theta u^k Y_k u^l Y_l = u^k u^l \left[\mathbb{F}_\theta c_k c_l - \left(\mathbb{F}_\theta c_k\right)\left(\mathbb{F}_\theta c_l\right)\right].$$

The second factor of the r.h.s. equals

$$\mathbb{F}_\theta v^k X_k v^l X_l = v^k v^l g_{kl}(\theta).$$

This proves (2.5).

Assume now that (2.6) holds. Combining it with the regularity condition (2.4) shows that c is an estimator for the escort family, with scaling function G. This makes it possible to write (2.6) as

$$\frac{1}{Z(\theta)P_\theta(x)}\frac{\partial}{\partial\theta^k}p_\theta(x) = \mathbb{F}_\theta c_k - c_k(x). \tag{2.7}$$

In this way one obtains

$$u^k u^l \left[\mathbb{F}_\theta c_k c_l - \left(\mathbb{F}_\theta c_k\right)\left(\mathbb{F}_\theta c_l\right)\right] = \frac{u^k u^l g_{kl}(\theta)}{Z(\theta)^2}. \tag{2.8}$$

On the other hand we have

$$\begin{aligned}\frac{\partial^2}{\partial\theta^l\partial\theta^k}F(\theta) &= \frac{\partial}{\partial\theta^l}\mathbb{E}_\theta c_k \\ &= \int_\Omega \mathrm{d}\mu(x)\,\frac{\partial p_\theta}{\partial\theta^l}(x)c_k(x) \\ &= Z(\theta)\int_\Omega \mathrm{d}\mu(x)\,P_\theta(x)c_k(x)\frac{\partial}{\partial\theta^k}\left[G(\theta)-\theta^l c_l(x)\right] \\ &= -Z(\theta)\left[\mathbb{F}_\theta c_k c_l - \left(\mathbb{F}_\theta c_k\right)\left(\mathbb{F}_\theta c_l\right)\right]. \qquad (2.9)\end{aligned}$$

Together with (2.8) this shows equality in (2.5) whenever $u = v$. □

It has not been investigated whether (2.6) is a necessary condition. For practical application of the lower bound one has to assume that c is also an estimator of the escort family $(P_\theta)_{\theta\in D}$, with scale function G. The previous proposition shows that this is automatically the case when (2.6) is satisfied.

Example 1. Let μ be the Lebesgue measure restricted to $[0,+\infty)$ and let

$$p_\theta(x) = \frac{2}{\theta}\left[1-\frac{x}{\theta}\right]_+ \tag{2.10}$$

with $\theta > 0$ and $[u]_+ = \max\{u, 0\}$. The Fisher information $I(\theta)$ is divergent. Hence, the usual lower bound of Cramér and Rao is useless.

Consider now the escort family

$$P_\theta(x) = \frac{1}{\theta}e^{-x/\theta}. \tag{2.11}$$

Then one calculates

$$g(\theta) = \frac{4}{\theta^2}(5e-13). \tag{2.12}$$

This fixes the r.h.s. of the inequality (2.5).

Let us estimate θ via its first moment, with $c(x) = 3x$. One has $\mathbb{E}_\theta c = \theta$, $\mathbb{E}_\theta c^2 = (3/2)\theta^2$, $F(\theta) = \theta^2/2$, $\mathbb{F}c = 3\theta$ and $\mathbb{F}c^2 = 18\theta^2$. Then (2.5) boils down to

$$\mathbb{F}c^2 - \big(\mathbb{F}c\big)^2 = 9\theta^2 \geq \frac{1}{4(5e-13)}\theta^2 \simeq 0.4\,\theta^2. \tag{2.13}$$

3. STATISTICAL MANIFOLD

The well-known example of a family with optimal estimator is the exponential family

$$p_\theta(x) = \exp\big(G(\theta) - \theta^k c_k(x)\big) \tag{3.1}$$

with

$$G(\theta) = -\log \int_\Omega \mathrm{d}\mu(x)\, e^{-\theta^k c_k(x)}. \tag{3.2}$$

One sees immediately that

$$\frac{\partial}{\partial\theta^k} p_\theta(x) = p_\theta(x)\left(\frac{\partial}{\partial\theta^k} G(\theta) - c_k(x)\right), \tag{3.3}$$

which is (2.7) with $Z(\theta)$ identically 1 and the escort pdf P_θ equal to p_θ. This example motivates also the geometric interpretation of (2.6), in the form (2.7), as a linear map between tangent planes. The score variables $\partial \log p_\theta/\partial\theta^k$ of the standard statistical manifold are replaced by the variables

$$\frac{1}{P_\theta(x)}\frac{\partial}{\partial\theta^k} p_\theta(x). \tag{3.4}$$

They are tangent vectors of the concave function $G(\theta) - \theta^l c_l$. The metric tensor of the latter function is a constant random variable. The geometry of the manifold of random variables $\big(G(\theta) - \theta^l c_l\big)_{\theta\in D}$ is transferred onto the family of pdfs $\big(p_\theta\big)_{\theta\in D}$.

Note that the score variables have vanishing expectation $\mathbb{F}_\theta$. It is now obvious to define an inner product of random variables by

$$\langle A, B\rangle_\theta = \mathbb{F}_\theta AB.$$

Then one has

$$\left\langle \frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^k}, \frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^l}\right\rangle_\theta = g_{kl}(\theta).$$

Let $g^{kl}(\theta)$ denote the inverse of $g_{kl}(\theta)$ (assume it exists). Then a projection operator π_θ onto the orthogonal complement of the tangent plane is defined by

$$\pi_\theta A = A - g^{kl}\left\langle \frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^k}, A\right\rangle_\theta \frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^l} - \mathbb{F}_\theta A.$$

If (2.6) is satisfied, then

$$\begin{aligned}
\pi_\theta \frac{\partial}{\partial\theta^l}\frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^k} &= \pi_\theta\left[\frac{\partial Z}{\partial\theta^l}\left(\mathbb{F}_\theta c_k - c_k\right) + Z(\theta)\frac{\partial^2 G}{\partial\theta^k\partial\theta^l}\right] \\
&= \frac{\partial Z}{\partial\theta^l}\left[\mathbb{F}_\theta c_k - c_k + g^{lm}(\theta)\langle\frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^l}, c_k\rangle_\theta \frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^m}\right] \\
&= \frac{\partial Z}{\partial\theta^l}\left[\mathbb{F}_\theta c_k - c_k - \frac{1}{Z(\theta)P_\theta}\frac{\partial p_\theta}{\partial\theta^k}\right] \\
&= 0.
\end{aligned}$$

This follows also immediately from

$$\frac{\partial}{\partial\theta^l}\frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^k}=\frac{1}{Z(\theta)}\frac{\partial Z}{\partial\theta^l}\frac{1}{P_\theta}\frac{\partial p_\theta}{\partial\theta^k}+Z(\theta)\frac{\partial^2 G}{\partial\theta^k\theta^l}.$$

That the derivatives of the score variables are linear combinations of the score variables and the constant random variable is usually the characteristic feature of the exponential family. This is a motivation to introduce a generalized notion of exponential family.

4. ϕ-LOGARITHMS AND ϕ-EXPONENTIALS

In the next section the notion of exponential family is generalized to a rather large class of families of pdfs. This is done by replacing the exponential function by some other function satisfying a minimal number of requirements. The latter function will be called a deformed exponential and will be denoted $\exp_\phi$. This has the advantage that the resulting expressions look very familiar, resembling those of the exponential family.

Fix an increasing function ϕ of $[0,+\infty)$, strictly positive on $(0,+\infty)$. It is used to define the ϕ-logarithm $\ln_\phi$ by

$$\ln_\phi(u)=\int_1^u \mathrm{d}v\,\frac{1}{\phi(v)},\qquad u>0. \tag{4.1}$$

Clearly, $\ln_\phi$ is a concave function which is negative on $(0,1)$ and positive on $(1,+\infty)$. The inverse of the function $\ln_\phi$ is denoted $\exp_\phi$. It is defined on the range of $\ln_\phi$. The definition can be extended to all of $\mathbb{R}$ by putting $\exp_\phi(u)=0$ if u is too small and $\exp_\phi=+\infty$ if u is too large. In case $\phi(u)=u$ for all u then $\ln_\phi$ coincides with the natural logarithm and $\exp_\phi$ coincides with the exponential function.

Given ϕ, introduce a function ψ of $\mathbb{R}$ by

$$\begin{aligned}\psi(u)&=\phi\big(\exp_\phi(u)\big) && \text{if } u \text{ is in the range of } \ln_\phi\\ &=0 && \text{if } u \text{ is too small}\\ &=+\infty && \text{if } u \text{ is too large.}\end{aligned} \tag{4.2}$$

Clearly is $\phi(u)=\psi(\ln_\phi(u))$ for all $u>0$.

Proposition 4.1. *One has for all u in $\mathbb{R}$*

$$\begin{aligned}0\le\exp_\phi(u)&=1+\int_0^u \mathrm{d}v\psi(v)\\ &=\int_{-\infty}^u \mathrm{d}v\psi(v)\le+\infty.\end{aligned} \tag{4.3}$$

Proof. First consider the case that $[0,u)$ belongs to the range of $\ln_\phi$. Then a substitution of integration variables $v=\ln_\phi(w)$ is possible. One finds, using $\mathrm{d}v/\mathrm{d}w=1/\phi(w)$ and $\psi(v)=\phi\big(\exp_\phi(v)\big)=\phi(w)$,

$$\begin{aligned}\int_0^u \mathrm{d}v\psi(v)&=\int_1^{\exp_\phi(u)}\mathrm{d}w\\ &=\exp_\phi(u)-1.\end{aligned}$$

Using $\exp_\phi(-\infty)=0$ one concludes (4.3).

In case $M = \sup_v \ln_\phi(v)$ is finite and $u \geq M$ then $\psi(v) = +\infty$ for $v \in [M, u]$. One has

$$\begin{aligned}\int_0^u \mathrm{d}v\psi(v) &\geq \int_0^M \mathrm{d}v\psi(v)\\ &= \int_1^{+\infty} \mathrm{d}w\\ &= +\infty.\end{aligned}$$

But also the l.h.s. of (4.3) is infinite. Hence the equality holds.

Finally, if $m = \inf_v \ln_\phi(v)$ is finite and $u \leq m$ then $\psi(v) = 0$ holds for $v \leq m$. Hence

$$\int_0^u \mathrm{d}v\,\psi(v) = \int_0^m \mathrm{d}v\,\psi(v) = \int_1^0 \mathrm{d}w = -1.$$

This ends the proof. □

Proposition 4.2. *The function $\exp_\phi$ is continuous on the open interval of points where it does not diverge.*

Proof. Let m and M be as in the proof of the previous proposition. Then $\exp_\phi$ is differentiable on (m, M). If $m = -\infty$ this ends the proof. If m is finite then it suffices to verify that $\exp_\phi(u)$ is continuous in $u = m$. But this is straightforward. □

Example 2. Let $\phi(u) = u^q$ with $q > 0$. This function is increasing and strictly positive on $(0, +\infty)$. Hence, it defines a ϕ-logarithm which will be denoted $\ln_q$ and is given by

$$\begin{aligned}\ln_q(u) &= \int_1^u \mathrm{d}v\,\frac{1}{v^q}\\ &= \frac{u^{1-q}-1}{1-q} \quad \text{if } q \neq 1\\ &= \log(u) \quad \text{if } q = 1.\end{aligned}$$

This deformed logarithm has been introduced in the context of nonextensive statistical physics in [18]. The inverse function is denoted $\exp_q$ and is given by

$$\exp_q(u) = \big[1 + (1-q)u\big]_+^{1/(1-q)}.$$

The function ψ is then given by

$$\psi(u) = \big[1 + (1-q)u\big]_+^{q/(1-q)}.$$

Example 3. Let $\phi(x) = \lceil x \rceil$, the smallest integer not smaller than x. This piecewise constant function is increasing and strictly positive on $(0, +\infty)$. Hence, $\ln_\phi$ is piecewise linear. The function ψ is given by

$$\begin{aligned}\psi(x) &= 0 \quad \text{if } x \leq -1\\ &= \phi(1+x) \quad \text{otherwise.}\end{aligned}$$

The ϕ-exponential $\exp_\phi$ is also piecewise linear and satisfies

$$\exp_\phi(x) = 0 \quad \text{if } x \leq -1.$$

5. THE ϕ-EXPONENTIAL FAMILY

Let ϕ be given as in the previous section. Fix a measure space Ω, μ and a set of random variables c_k, $k = 1, \ldots, n$. The ϕ-exponential family of pdfs $\big(p_\theta\big)_{\theta\in D}$ is defined by

$$p_\theta(x) = \exp_\phi\big(G(\theta) - \theta^k c_k(x)\big). \tag{5.1}$$

The domain D is an open set of θ for which $G(\theta)$ exists such that (5.1) is properly normalized, i.e. $p_\theta \in \mathcal{M}_1(\mu)$. The distributions (5.1) are the equilibrium pdfs of generalized thermostatistics as introduced in [11, 12].

Proposition 5.1. *The function $G(\theta)$ is concave on D.*

Proof. Assume θ, η and $\lambda\theta + (1-\lambda)\eta$ in D for some λ in $[0,1]$. Then, using the convexity of $\exp_\phi$,

$$\exp_\phi\big(\lambda G(\theta) + (1-\lambda)G(\eta) - \big[\lambda\theta^k + (1-\lambda)\eta^k\big]c_k(x)\big) \le \lambda p_\theta(x) + (1-\lambda)p_\eta(x).$$

Hence

$$\int_{\mathbb{R}^n} \mathrm{d}\mu(x)\, \exp_\phi\big(\lambda G(\theta) + (1-\lambda)G(\eta) - \big[\lambda\theta^k + (1-\lambda)\eta^k\big]c_k(x)\big) \le 1.$$

Compare this with

$$\int_{\mathbb{R}^n} \mathrm{d}\mu(x)\, \exp_\phi\big(G(\lambda\theta + (1-\lambda)\eta) - \big[\lambda\theta^k + (1-\lambda)\eta^k\big]c_k(x)\big) = 1.$$

Since $\exp_\phi$ is increasing one concludes that

$$\lambda G(\theta) + (1-\lambda)G(\eta) \le G(\lambda\theta + (1-\lambda)\eta).$$

This means that G is concave. □

Proposition 5.2. *Let ψ be determined by ϕ via (4.2). If the integral*

$$Z(\theta) = \int_\Omega \mathrm{d}\mu(x)\, \psi\big(G(\theta) - \theta^k c_k(x)\big)$$

converges for all $\theta \in D$, then $\big(p_\theta\big)_{\theta\in D}$ has an escort family $\big(P_\theta\big)_{\theta\in D}$, given by

$$\begin{aligned} P_\theta(x) &= \frac{1}{Z(\theta)}\phi\big(p_\theta(x)\big) && \text{if } p_\theta(x) > 0 \\ &= 0 && \text{otherwise}. \end{aligned}$$

Condition (2.6) is satisfied.

Proof. One has

$$\begin{aligned} \phi\big(p_\theta(x)\big) &= \phi\big(\exp_\phi\big(G(\theta) - \theta^k c_k(x)\big)\big) \\ &= \psi\big(G(\theta) - \theta^k c_k(x)\big). \end{aligned}$$

Because $\phi\big(p_\theta(x)\big)$ cannot be zero for μ-almost all x one concludes that $Z(\theta) > 0$ and that P_θ is properly normalized.

From the properties of the function $\exp_\phi$ follows immediately that

$$\begin{aligned} \frac{\partial}{\partial\theta^l} p_\theta(x) &= \psi\big(G(\theta) - \theta^k c_k(x)\big)\frac{\partial}{\partial\theta^l}\big(G(\theta) - \theta^m c_m(x)\big) \\ &= Z(\theta)P_\theta(x)\frac{\partial}{\partial\theta^l}\big(G(\theta) - \theta^m c_m(x)\big). \end{aligned}$$

This proves that $\big(P_\theta\big)_{\theta\in D}$ satisfies (2.6). □

Example 2 continued. Let $\phi(u) = u^q$ as in Example 2 above. The pdfs p_θ are given by

$$p_\theta(x) = \big[1 + (1-q)\big(G(\theta) - \theta^k c_k(x)\big)\big]_+^{1/(1-q)}, \tag{5.2}$$

for θ in a suitable domain D. The escort probabilities are

$$P_\theta(x) = \frac{1}{Z(\theta)} \big[1 + (1-q)\big(G(\theta) - \theta^k c_k(x)\big)\big]_+^{q/(1-q)} \tag{5.3}$$

with

$$Z(\theta) = \int_\Omega \mathrm{d}\mu(x)\ \big[1 + (1-q)\big(G(\theta) - \theta^k c_k(x)\big)\big]_+^{q/(1-q)}$$

(assuming convergence of these integrals). The family $\big(p_\theta\big)_{\theta\in D}$ coincides with Amari's α-family [3], with α given by $\alpha = 2q - 1$.

Example 1 continued. Example 1 is the $q = 0$-limit of Example 2. Let $\phi(u) = 1$ for all $u > 0$. Then

$$\begin{aligned} \ln_\phi(u) &= u - 1 \\ \exp_\phi(u) &= [1+u]_+ \\ \psi(u) &= 1 \qquad \text{if } u > -1; \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

One has

$$p_\theta(x) = \frac{2}{\theta}\left[1 - \frac{x}{\theta}\right]_+ = \exp_\phi\left(\frac{2}{\theta} - 1 - \frac{2x}{\theta^2}\right).$$

This is a ϕ-exponential family with parameter $\Theta = 1/\theta^2$, estimator $c(x) = 2x$ and scale function $G(\Theta) = 2\sqrt{\Theta}$. The escort probabilities, making inequality (2.5) optimally satisfied, are given by

$$P_\Theta(x) = \frac{1}{\theta}\mathbb{I}_{0\le x\le\theta}.$$

The information matrix $g(\Theta)$ equals $\theta^4/3$. Further is $\mathbb{F}_\Theta c = \theta$ and $\mathbb{F}_\Theta c^2 = 4\theta^2/3$ and

$$\frac{\partial}{\partial\Theta}F(\Theta) = \mathbb{E}_\Theta c = 2\theta/3 = 2/3\sqrt{\Theta}.$$

It is now straightforward to verify that the inequality (2.5) is optimally satisfied.

6. DIVERGENCES

Divergences of the Bregman type are needed for what follows. In the form given below they have been introduced in [9].

Fix a strictly positive increasing function ϕ of $[0, +\infty)$. Introduce

$$D_\phi(p||p') = \int_\Omega \mathrm{d}\mu(x) \int_{p'(x)}^{p(x)} \mathrm{d}u\ \left[\ln_\phi(u) - \ln_\phi(p'(x))\right]. \tag{6.1}$$

$D_\phi(p||p') \ge 0$ follows because $\ln_\phi$ is an increasing function. Also convexity in the first argument follows because $\ln_\phi$ is an increasing function.

Let $\big(p_\theta\big)_{\theta\in D}$ be ϕ-exponential. Then infinitesimal variation of the divergence $D_\phi(p||p')$ reproduces the metric tensor $g_{kl}(\theta)$, up to a scalar function. Indeed, one has

$$\begin{aligned} \frac{\partial}{\partial\theta^k} D_\phi(p_\theta||p_\eta)\big|_{\eta=\theta} &= 0 \\ \frac{\partial}{\partial\eta^k} D_\phi(p_\theta||p_\eta)\big|_{\eta=\theta} &= 0 \end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial\theta^k\partial\theta^l}D_\phi(p_\theta||p_\eta)\Big|_{\eta=\theta} &= \frac{\partial}{\partial\theta^k}\int_\Omega \mathrm{d}\mu(x)\ \left[\ln_\phi\left(p_\theta(x)\right)-\ln_\phi\left(p_\eta(x)\right)\right]\frac{\partial}{\partial\theta^l}p_\theta(x)\Big|_{\eta=\theta}\\ &= \int_\Omega \mathrm{d}\mu(x)\,\frac{1}{\phi\big(p_\theta(x)\big)}\left[\frac{\partial}{\partial\theta^k}p_\theta(x)\right]\left[\frac{\partial}{\partial\theta^l}p_\theta(x)\right]\\ &= \frac{1}{Z(\theta)}g_{kl}(\theta).\end{aligned}$$

Similar calculations give

$$-\frac{\partial^2}{\partial\theta^k\partial\eta^l}D_\phi(p_\theta||p_\eta)\Big|_{\eta=\theta} = \frac{\partial^2}{\partial\eta^k\partial\eta^l}D_\phi(p_\theta||p_\eta)\Big|_{\eta=\theta} = \frac{1}{Z(\theta)}g_{kl}(\theta).$$

7. INFORMATION CONTENT

In [10] the definition of a deformed logarithm contains the additional condition that the integral

$$\int_1^0 \mathrm{d}u\ \ln_\phi(u) = \int_0^1 \mathrm{d}u\,\frac{u}{\phi(u)} < +\infty$$

converges. This condition is needed in the definition of entropy functional / information content based on the deformed logarithm. Introduce another strictly increasing positive function χ by

$$\chi(v) = \left[\int_0^{1/v}\mathrm{d}u\,\frac{u}{\phi(u)}\right]^{-1}$$

The motivation for introducing this function comes from the fact that it satisfies the following property.

Lemma 7.1.

$$\frac{\mathrm{d}}{\mathrm{d}v}v\ln_\chi(1/v) = -\ln_\phi(v) - \int_0^1 \mathrm{d}u\,\frac{u}{\phi(u)}. \tag{7.1}$$

Proof.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}v}v\ln_\chi(1/v) &= \ln_\chi(1/v) - \frac{1}{v\chi(1/v)}\\ &= \int_1^{1/v}\mathrm{d}u\,\frac{1}{\chi(u)} - \frac{1}{v}\int_0^v \mathrm{d}u\,\frac{u}{\phi(u)}\\ &= \int_1^{1/v}\mathrm{d}u\int_0^{1/u}\mathrm{d}z\,\frac{z}{\phi(z)} - \frac{1}{v}\int_0^v \mathrm{d}u\,\frac{u}{\phi(u)}\\ &= -\int_1^v \mathrm{d}u\,\frac{1}{u^2}\int_0^u \mathrm{d}z\,\frac{z}{\phi(z)} - \frac{1}{v}\int_0^v \mathrm{d}u\,\frac{u}{\phi(u)}\\ &= -\int_0^1 \mathrm{d}z\,\frac{z}{\phi(z)} - \ln_\phi(v),\end{aligned}$$

which is the desired result. □

Define the information content (also called entropy functional) $I_\phi(p)$ of a pdf p in $\mathcal{M}_1(\mu)$ by

$$I_\phi(p) = \int_\Omega \mathrm{d}\mu(x)\,p(x)\ln_\chi(1/p(x))$$

whenever the integral converges. Using the lemma one verifies immediately that $I_\phi(p)$ is a concave function of p. A short calculation gives

$$\begin{aligned} I_\phi(p) &= \int_\Omega \mathrm{d}\mu(x)\, p(x) \int_1^{1/p(x)} \mathrm{d}u\, \frac{1}{\chi(u)} \\ &= \int_\Omega \mathrm{d}\mu(x)\, p(x) \int_1^{p(x)} \frac{1}{\chi(1/v)}\, \mathrm{d}\frac{1}{v} \\ &= \int_\Omega \mathrm{d}\mu(x)\, p(x) \int_1^{p(x)} \left[\int_0^v \mathrm{d}u\, \frac{u}{\phi(u)}\right] \mathrm{d}\frac{1}{v} \\ &= \int_\Omega \mathrm{d}\mu(x)\, p(x) \left[\frac{1}{p(x)\chi\big(1/p(x)\big)} - \frac{1}{\chi(1)} - \ln_\phi\big(p(x)\big)\right] \\ &= -\frac{1}{\chi(1)} - \int_\Omega \mathrm{d}\mu(x) \int_0^{p(x)} \mathrm{d}u\, \ln_\phi(u). \end{aligned}$$

This implies that

$$I_\phi(p) - I_\phi(p') = -\int_\Omega \mathrm{d}\mu(x) \int_{p'(x)}^{p(x)} \mathrm{d}u\, \ln_\phi(u),$$

and hence

$$D_\phi(p||p') = I_\phi(p') - I_\phi(p) - \int_\Omega \mathrm{d}\mu(x)\, \big(p(x) - p'(x)\big) \ln_\phi\big(p'(x)\big). \tag{7.2}$$

This relation links the divergence $D_\phi(p||p')$ with the information function $I_\phi(p)$.

The following result shows that the ϕ-exponential family is a conditional maximizer of I_ϕ. It also shows that the scale function F is the Legendre transform of the information content I_ϕ

Theorem 7.2. *Let $\big(p_\theta\big)_{\theta\in D}$ be ϕ-exponential, with estimator c and scale functions F and G. Then there exists a constant F_0 such that*

$$F(\theta) = F_0 + \min_{p\in\mathcal{M}_1(\mu)} \{\mathbb{E}_p\theta^k c_k - I_\phi(p)\}. \tag{7.3}$$

The minimum is attained for $p = p_\theta$. In particular, $F(\theta)$ is a concave function of θ and p_θ maximizes $I_\phi(p)$ under the constraint that

$$\mathbb{E}_p\theta^k c_k = \mathbb{E}_\theta\theta^k c_k.$$

Proof. Let us first show that for any pdf p

$$\mathbb{E}_p\theta^k c_k - I_\phi(p) \geq \mathbb{E}_\theta\theta^k c_k - I_\phi(p_\theta). \tag{7.4}$$

One has

$$\begin{aligned} \int_\Omega \mathrm{d}\mu(x)\, \big(p(x) - p_\theta(x)\big) \ln_\phi\big(p_\theta(x)\big) &= \int_\Omega \mathrm{d}\mu(x)\, \big(p(x) - p_\theta(x)\big) \big[G(\theta) - \theta^k c_k\big] \\ &= -(\mathbb{E}_p - \mathbb{E}_\theta)\theta^k c_k. \end{aligned}$$

Hence, (7.2) becomes now

$$D_\phi(p||p_\theta) = I_\phi(p_\theta) - I_\phi(p) + (\mathbb{E}_p - \mathbb{E}_\theta)\theta^k c_k.$$

But one has always $D_\phi(p||p_\theta) \geq 0$. Therefore, (7.4) follows.

Next calculate, using the lemma,

$$\begin{aligned}\frac{\partial}{\partial\theta^k}I_\phi(p_\theta) &= \int \mathrm{d}\mu(x)\,\left(-\ln_\phi\big(p_\theta(x)\big) - \int_0^1 \mathrm{d}u\,\frac{u}{\phi(u)}\right)\frac{\partial}{\partial\theta^k}p_\theta(x)\\ &= \int \mathrm{d}\mu(x)\,\left(-G(\theta)+\theta^l c_l(x) - \int_0^1 \mathrm{d}u\,\frac{u}{\phi(u)}\right)\frac{\partial}{\partial\theta^k}p_\theta(x)\\ &= \int \mathrm{d}\mu(x)\,\big(\theta^l c_l(x)\big)\,\frac{\partial}{\partial\theta^k}p_\theta(x)\\ &= \frac{\partial}{\partial\theta^k}\big(\mathbb{E}_\theta\theta^l c_l\big) - \mathbb{E}_\theta c_k.\end{aligned}$$

Because c is an estimator with scale function F one obtains

$$\frac{\partial}{\partial\theta^k}\big(\mathbb{E}_\theta\theta^l c_l - I_\phi(p_\theta)\big) = \frac{\partial}{\partial\theta^k}F(\theta).$$

Hence there exists a constant F_0 for which

$$F(\theta) = F_0 + \mathbb{E}_\theta\theta^l c_l - I_\phi(p_\theta). \tag{7.5}$$

In combination with (7.4) this results in (7.3). □

Without restriction one can assume $F_0 = 0$. In statistical physics the function $F(\theta)$ is free energy divided by temperature.

Example 4. Let $\phi(u) = u^{2-q}/q$, with $0 < q < 2$. This is of course only a re-parametrization of Example 2, which is done to recover expressions found in the literature. The deformed logarithm is given by

$$\begin{aligned}\ln_\phi(u) &= \frac{q}{q-1}(u^{q-1}-1) \qquad &\text{if } q\neq 1\\ &= \log(u) &\text{if } q = 1.\end{aligned}$$

One obtains $\chi(v) = v^q$ and hence

$$I_\phi(p) = \int \mathrm{d}\mu(x)\,p(x)\frac{1-p(x)^{q-1}}{q-1}. \tag{7.6}$$

This is the entropy functional proposed by Tsallis [17] as a basis for nonextensive thermostatistics, and reported earlier in the literature by Havrda and Charvat [7] and by Daróczy [5]. The corresponding expression for the divergence is

$$D_\phi(p||p') = \frac{1}{q-1}\int \mathrm{d}\mu(x)\,p(x)\left[p(x)^{q-1} - p'(x)^{q-1}\right] - \int \mathrm{d}\mu(x)\,\left[p(x)-p'(x)\right]p'(x)^{q-1}.$$

By Theorem 7.2, the pdf p_θ minimizes 'free energy' $\mathbb{E}_p\theta^k c_k - I_\phi(p)$. But note that, due to the re-parametrization, p_θ is not given by (5.2), but equals

$$p_\theta = \big[1+(1-q')\big(G(\theta)-\theta^k c_k(x)\big)\big]_+^{q'/(1-q')}$$

with $q' = 1/q$. The latter expression coincides with that of the escort pdf (5.3), with q replaced by q' and with incorporation of the normalization $Z(\theta)$ into the scale function $G(\theta)$. The Tsallis literature [19] associates with each pdf p an escort pdf P by the relation $P \sim p^q$. Then, expression (7.6) is optimized under the constraint that $\mathbb{E}_P c_k$ have given values. The resulting formalism differs slightly from the present one.

8. DUAL COORDINATES

Introduce dual coordinates

$$\eta_k = \mathbb{E}_\theta c_k = \frac{\partial F}{\partial \theta^k}. \tag{8.1}$$

Assume (2.6) holds. Then, one obtains from (2.9)

$$\begin{aligned}\frac{\partial \eta_k}{\partial \theta^l} &= \frac{\partial}{\partial \theta^l}\mathbb{E}_\theta c_k \\ &= \frac{\partial^2}{\partial \theta^l \partial \theta^k} F(\theta) \\ &= -Z(\theta)\left[\mathbb{F}_\theta c_k c_l - \left(\mathbb{F}_\theta c_k\right)\left(\mathbb{F}_\theta c_l\right)\right] \\ &= -\frac{1}{Z(\theta)} g_{kl}(\theta).\end{aligned}$$

To obtain the last line a ϕ-exponential family has been assumed. This relation implies

$$\frac{\partial \theta^k}{\partial \eta_l} = -Z(\theta) g^{kl}(\theta). \tag{8.2}$$

These are the orthogonality relations between the two sets of coordinates θ and η. Next we derive the dual relation of (8.1).

Proposition 8.1. *Let $\left(p_\theta\right)_{\theta \in D}$ be ϕ-exponential. Assume the regularity condition (2.4) is satisfied. Then*

$$\theta^k = \frac{\partial}{\partial \eta_k} I_\phi(p_\theta). \tag{8.3}$$

Proof. One calculates (assume integration and partial derivative can be interchanged), using Lemma 7.1,

$$\begin{aligned}\frac{\partial}{\partial \theta^k} I_\phi(p_\theta) &= -\int_\Omega \mathrm{d}\mu(x)\, \left[\ln_\phi\left(p_\theta(x)\right) + \int_0^1 \mathrm{d}u\, \frac{u}{\phi(u)}\right] \frac{\partial}{\partial \theta^k} p_\theta(x) \\ &= -\int_\Omega \mathrm{d}\mu(x)\, \left[G(\theta) - \theta^l c_l(x) + \int_0^1 \mathrm{d}u\, \frac{u}{\phi(u)}\right] \frac{\partial}{\partial \theta^k} p_\theta(x) \\ &= \int_\Omega \mathrm{d}\mu(x)\, \theta^l c_l(x) \frac{\partial}{\partial \theta^k} p_\theta(x).\end{aligned}$$

To obtain the last line the regularity condition has been used. Use now that p_θ satisfies (2.6). One obtains

$$\begin{aligned}\frac{\partial}{\partial \theta^k} I_\phi(p_\theta) &= Z(\theta)\mathbb{F}_\theta \theta^l c_l (\mathbb{F}_\theta c_k - c_k) \\ &= -Z(\theta)\theta^l g_{lk}(\theta).\end{aligned}$$

In combination with (8.2) this gives

$$\begin{aligned}\frac{\partial}{\partial \eta_l} I_\phi(p_\theta) &= \left(\frac{\partial}{\partial \theta^l} I_\phi(p_\theta)\right) \frac{\partial \theta^k}{\partial \eta_l} \\ &= \left(-Z(\theta)\theta^m g_{ml}(\theta)\right)\left(-\frac{1}{Z(\theta)} g^{kl}(\theta)\right) \\ &= \theta^l.\end{aligned}$$

□

Equation (8.3) is the dual relation of (8.1). Expression (7.5) can now be written as

$$F(\theta) + E(\eta) = \theta^k \eta_k \tag{8.4}$$

with $E(\eta) = I_\phi(p_\theta)$.

9. DISCUSSION

The present paper introduces generalized exponential families, and calls them ϕ-exponential because they depend on the choice of a strictly positive non-decreasing function ϕ of $(0, +\infty)$. Several properties, known to hold for the exponential family, can be generalized. The paper starts with a generalization of the well-known lower bound of Cramér and Rao, involving the concept of escort probability distributions. See Theorem 2.1. It is shown that the ϕ-exponential family optimizes this generalized lower bound. The metric tensor, which generalizes the Fisher information, depends on both the family of pdfs and the escort family, and determines the geometry of the statistical manifold.

The final part of the paper deals with the dual structure of the statistical manifold, which survives in the more general context of ϕ-exponential families. It is shown in Theorem 7.2 that the ϕ-exponential family satisfies a variational principle with respect to a suitably defined entropy functional. The well-known duality of statistical physics, between energy and temperature and between entropy and free energy, is recovered.

Throughout the paper the number of parameters n has been assumed to be finite. A non-parametrized approach to statistical manifolds is found in [13]. The extension of the present work to this more abstract context has not been considered.

REFERENCES

[1] S. ABE, q-Deformed Entropies and Fisher Metrics, in: *Proceedings of The 5th International Wigner Symposium*, (August 25-29, 1997, Vienna, Austria), eds. P. Kasperkovitz and D. Grau (World Scientific, Singapore, 1998) p. 66.

[2] S. ABE, Geometry of escort distributions, arXiv:cond-mat/0305231, *Phys. Rev.*, E**68** (2003), 031101.

[3] S. AMARI, Differential-geometrical methods in statistics, *Lecture Notes in Statistics*, **28** (1985).

[4] C. BECK AND F. SCHLÖGL, *Thermodynamics of chaotic systems: An introduction*, Cambridge University Press, Cambridge, 1993.

[5] Z. DARÓCZY, *Inform. Control*, **16** (1970), 36– .

[6] J.W. GIBBS, *Elementary principles in statistical mechanics developed with special reference to the rational foundation of thermodynamics*, Dover, 1960.

[7] J. HAVRDA AND F. CHARVAT, *Kybernetica*, **3** (1967), 30–35.

[8] M.K. MURRAY AND J.W. RICE, *Differential geometry and statistics*, Chapman and Hall, 1993.

[9] J. NAUDTS, Continuity of a class of entropies and relative entropies, arXiv:cond-mat/0208038, *Rev. Math. Phys.*, **16**(6) (2004), 809–822.

[10] J. NAUDTS, Deformed exponentials and logarithms in generalized thermostatistics, arXiv:cond-mat/0203489, *Physica*, **A**316 (2002), 323–334.

[11] J. NAUDTS, Generalized thermostatistics and mean-field theory, arXiv:cond-mat/0211444, *Physica*, A**332** (2004), 279–300.

[12] J. NAUDTS, Generalized thermostatistics based on deformed exponential and logarithmic functions, arXiv:cond-mat/0311438, *Physica*, A**340** (2004), 32–40.

[13] G. PISTONE AND C. SEMPI, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Ann. Statist*., **23** (1995), 1543–1561.

[14] M. SHIINO, H-theorem with generalized relative entropies and the Tsallis statistics, *J. Phys. Soc. Jpn.*, **67**(11) (1998), 3658–3660.

[15] R. TRASARTI-BATTISTONI, Euclidean and Riemannian geometrical approaches to non-extensive thermo-statistical mechanics, arXiv:cond-mat/0203536.

[16] C. TSALLIS, Generalized entropy-based criterion for consistent testing, *Phys. Rev*., E**58**(2) (1998), 1442–1445.

[17] C. TSALLIS, Possible Generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys*., **52** (1988), 479–487.

[18] C. TSALLIS, What are the numbers that experiments provide?, *Quimica Nova,* **17** (1994), 468.

[19] C. TSALLIS, R.S. MENDES AND A.R. PLASTINO, The role of constraints within generalized non-extensive statistics, *Physica*, A**261** (1998), 543–554.