Covariance and Fisher information in quantum mechanics

2002 J. Phys. A: Math. Gen. 35 929

(http://iopscience.iop.org/0305-4470/35/4/305)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 929–939 PII: S0305-4470(02)30015-5

Covariance and Fisher information in quantum mechanics

Dénes Petz

Department for Mathematical Analysis, Budapest University of Technology and Economics, H-1521 Budapest XI, Hungary

Received 22 October 2001
Published 18 January 2002
Online at stacks.iop.org/JPhysA/35/929

Abstract

Variance and Fisher information are ingredients of the Cramér–Rao inequality. We regard Fisher information as a Riemannian metric on a quantum statistical manifold and choose monotonicity under coarse graining as the fundamental property of variance and Fisher information. In this approach we show that there is a kind of dual one-to-one correspondence between the candidates of the two concepts. We emphasize that Fisher information is obtained from relative entropies as contrast functions on the state space and argue that the scalar curvature might be interpreted as an uncertainty density on a statistical manifold.

PACS numbers: 02.50.-r, 02.40.Ky, 03.65.Ta, 05.30.-d

1. Introduction

On the one hand standard quantum mechanics is a statistical theory, on the other hand, there is a so-called geometrical approach to mathematical statistics [1, 5]. In this paper the two topics are combined and the concept of covariance and Fisher information is studied from an abstract point of view. We begin with the Cramér–Rao inequality to realize that the two concepts are very strongly related. What they have in common is a kind of monotonicity property under coarse grainings. (Formally the monotonicity of covariance slightly differs from that of Fisher information.) Monotone quantities of Fisher information type determine an operator $\mathbb{J}$ which immediately gives a kind of generalized covariance. In this way a one-to-one correspondence is established between the candidates of the two concepts. In the paper we prove a Cramér–Rao type inequality in the setting of generalized variance and Fisher information. Moreover, we argue that the scalar curvature of the Fisher information Riemannian metric has a statistical interpretation. This gives an interpretation of an earlier formulated, but still open, conjecture on the monotonicity of the scalar curvature.

2. The Cramér–Rao inequality as an introduction

The Cramér–Rao inequality belongs to the fundamentals of estimation theory in mathematical statistics. Its quantum analogue was discovered immediately after the foundation of mathematical quantum estimation theory in the 1960s, see Helstrom [14], or Holevo [15] for a rigorous summary of the subject. Although both the classical Cramér–Rao inequality and its quantum analogue are as trivial as the Schwarz inequality, the subject attracts much attention because it is located on the highly exciting boundary of statistics, information and quantum theory.

As a starting point we present a very general form of the quantum Cramér–Rao inequality in the simple setting of finite dimensional quantum mechanics. For $\theta \in (-\varepsilon, \varepsilon) \subset \mathbb{R}$ a statistical operator D_θ is given and the aim is to estimate the value of the parameter θ close to 0. Formally D_θ is an $n \times n$ positive semidefinite matrix of trace 1 which describes a mixed state of a quantum mechanical system and we assume that D_θ is smooth (in θ). In our approach we deal with mixed states contrary to several other authors, see [9], for example. Assume that an estimation is performed by the measurement of a self-adjoint matrix A playing the role of an observable. A is called a *locally unbiased estimator* if

$$\frac{\partial}{\partial\theta} \mathrm{Tr}\ D_\theta A\Big|_{\theta=0} = 1. \tag{1}$$

This condition holds if A is an *unbiased estimator* for θ, that is:

$$\mathrm{Tr}\ D_\theta A = \theta \qquad (\theta \in (-\varepsilon, \varepsilon)). \tag{2}$$

To require this equality for all values of the parameter is a serious restriction on the observable A and we prefer to use the weaker condition (1).

Let $\varphi_0[\,\cdot\,,\,\cdot\,]$ be an inner product on the linear space of self-adjoint matrices. $\varphi_0[\,\cdot\,,\,\cdot\,]$ depends on the density matrix D_0, the notation reflects this fact. When D_θ is smooth in θ, as assumed earlier, the correspondence

$$B \mapsto \frac{\partial}{\partial\theta} \mathrm{Tr}\ D_\theta B\Big|_{\theta=0} \tag{3}$$

is a linear functional on the self-adjoint matrices and of the form $\varphi_0[B, L]$ with some $L = L^*$. From (1) and (3) we have $\varphi_0[A, L] = 1$ and the Schwarz inequality yields

$$\varphi_0[A, A] \geqslant \frac{1}{\varphi_0[L, L]}. \tag{4}$$

This is the celebrated inequality of Cramér–Rao type for the locally unbiased estimator. We want to interpret the left-hand side as a *generalized variance* of A.

The right-hand side of (4) is independent of the estimator and provides a lower bound for the generalized variance. The denominator $\varphi_0[L, L]$ appears to be in the role of Fisher information here. We call it *quantum Fisher information* with respect to the generalized variance $\varphi_0[\,\cdot\,,\,\cdot\,]$. This quantity depends on the tangent of the curve D_θ.

We want to conclude from the earlier argument that whatever Fisher information and generalized variance are in the quantum mechanical setting, they are very strongly related. In an earlier work [21, 24] we used a monotonicity condition to make a limitation on the class of Riemannian metrics on the state space of a quantum system. The monotone metrics are called Fisher information quantities in this paper. Now we observe that a similar monotonicity property can be used to obtain a class of bilinear forms; we call the elements of this class generalized variances. The usual variance of two observables is included with many other quantities as well. We describe a one-to-one correspondence between variances and Fisher information. The correspondence is given by an operator $\mathbb{J}$ which immediately appears in the analysis of the inequality (4).

Since the sufficient and necessary condition for the equality in the Schwarz inequality is well-known, we are able to analyse the case of equality in (4). The condition for equality is

$$A = \lambda L$$

for some constant $\lambda \in \mathbb{R}$. On the $n \times n$ self-adjoint matrices we have two inner products: $\varphi_0[\,\cdot\,,\,\cdot\,]$ and $\langle A, B\rangle := \mathrm{Tr}\ AB$. There exists a linear operator $\mathbb{J}_0$ on the self-adjoint matrices such that

$$\varphi_0[A, B] = \mathrm{Tr}\ A\mathbb{J}_0(B).$$

Therefore the necessary and sufficient condition for equality in (4) is

$$\dot{D}_0 := \frac{\partial}{\partial\theta} D_\theta\Big|_{\theta=0} = \lambda^{-1}\mathbb{J}_0(A). \tag{5}$$

Therefore there exists a unique locally unbiased estimator $A = \lambda\mathbb{J}_0^{-1}(\dot{D}_0)$, where the number λ is chosen in such a way that condition (1) should be satisfied.

3. Coarse graining and Fisher information

In the simple setting in which the state is described by a density matrix, a coarse graining is an affine mapping sending density matrices into density matrices. Such a mapping extends to all matrices and provides a positivity and trace preserving linear transformation. A common example of coarse graining sends the density matrix D_{12} of a composite system $1+2$ into the (reduced) density matrix D_1 of component 1. There are several reasons to assume complete positivity—as we do here—about a coarse graining.

Assume that D_θ is a smooth curve of density matrices with tangent $A := \dot{D}_0$ at D_0. The quantum Fisher information $F_D(A)$ is an information quantity associated with the pair (D_0, A), it appeared in the Cramér–Rao inequality earlier and the Fisher information gives a bound for the (generalized) variance of a locally unbiased estimator. Now let α be a coarse graining. Then $\alpha(D_\theta)$ is another curve in the state space. Due to the linearity of α, the tangent at $\alpha(D_0)$ is $\alpha(A)$. As is usual in statistics, information cannot be gained by coarse graining, therefore we expect that the Fisher information at the density matrix D_0 in the direction A must be larger than the Fisher information at $\alpha(D_0)$ in the direction $\alpha(A)$. This is the *monotonicity property of the Fisher information* under coarse graining

$$F_D(A) \geqslant F_{\alpha(D)}(\alpha(A)). \tag{6}$$

Although we do not want to have a concrete formula for the quantum Fisher information, we require that this monotonicity condition must hold. Another requirement is that $F_D(A)$ should be quadratic in A, in other words there exists a non-degenerate real bilinear form $\gamma_D(A, B)$ on the self-adjoint matrices such that

$$F_D(A) = \gamma_D(A, A). \tag{7}$$

The requirements (6) and (7) are strong enough to obtain a reasonable but still wide class of possible quantum Fisher information.

We may assume that

$$\gamma_D(A, B) = \mathrm{Tr}\ A\mathbb{J}_D^{-1}(B^*), \tag{8}$$

for an operator $\mathbb{J}_D$ acting on matrices. (This formula expresses the inner product γ_D by means of the Hilbert–Schmidt inner product and the positive linear operator $\mathbb{J}_D$.) In terms of the operator $\mathbb{J}_D$ the monotonicity condition reads as

$$\alpha^*\mathbb{J}_{\alpha(D)}^{-1}\alpha \leqslant \mathbb{J}_D^{-1} \tag{9}$$

for every coarse graining α. (α^* stand for the adjoint of α with respect to the Hilbert–Schmidt product. Recall that α is completely positive and trace preserving if and only if α^* is completely positive and unital.) On the other hand the latter condition is equivalent to

$$\alpha \mathbb{J}_D \alpha^* \leqslant \mathbb{J}_{\alpha(D)}. \tag{10}$$

We proved the following theorem in [21], see also [17, 25].

Theorem 3.1. *If for every density matrix D a positive definite bilinear form γ_D is given such that (6) holds for all completely positive coarse grainings α and $\gamma_D(A, A)$ is continuous in D for every fixed A, then there exists a unique operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}$ such that $f(t) = tf(t^{-1})$ and $\gamma_D(A, A)$ is given by the following prescription*

$$\gamma_D(A, A) = \mathrm{Tr}\ A\mathbb{J}_D^{-1}(A) \qquad \textit{and} \qquad \mathbb{J}_D = \mathbb{R}_D^{1/2} f(\mathbb{L}_D \mathbb{R}_D^{-1}) \mathbb{R}_D^{1/2}$$

where the linear transformations $\mathbb{L}_D$ and $\mathbb{R}_D$ acting on matrices are the left and right multiplications; that is,

$$\mathbb{L}_D(X) = DX \qquad \textit{and} \qquad \mathbb{R}_D(X) = XD.$$

Although the statement of the theorem seems to be rather complicated, the formula for $F_D(A) = \gamma_D(A, A)$ becomes simpler when D and A commute. On the subspace $\{A : AD = DA\}$ the left multiplication $\mathbb{L}_D$ coincides with the right one $\mathbb{R}_D$ and $f(\mathbb{L}_D \mathbb{R}_D^{-1}) = f(1)$. Therefore we have

$$F_D(A) = \frac{1}{f(1)} \mathrm{Tr}\ D^{-1} A^2 \qquad \text{if} \quad AD = DA. \tag{11}$$

Under the hypothesis of commutation the quantum Fisher information is unique up to a constant factor. (This fact reminds us of the Cencov uniqueness theorem in the Kolmogorovian probability [5]. According to this theorem the metric on finite probability spaces is unique when monotonicity under Markovian kernels is posed.) We say that the quantum Fisher information is *classically Fisher-adjusted* if

$$F_D(A) = \mathrm{Tr}\ D^{-1} A^2 \qquad \text{when} \quad AD = DA. \tag{12}$$

This means that we impose the normalization $f(1) = 1$ on the operator monotone function. In the sequel we always assume this condition.

Via the operator $\mathbb{J}_D$, each monotone Fisher information determines a quantity

$$\varphi_D[A, A] := \mathrm{Tr}\ A\mathbb{J}_D(A) \tag{13}$$

which could be called *generalized variance*. According to (10) this possesses the monotonicity property

$$\varphi_D[\alpha^*(A), \alpha^*(A)] \leqslant \varphi_{\alpha(D)}[A, A]. \tag{14}$$

Since (9) and (10) are equivalent we observe a *one-to-one correspondence between monotone Fisher information and monotone generalized variances*. Any such variance has the property $\varphi_D[A, A] = \mathrm{Tr}\ DA^2$ for commuting D and A. The later examples show that it is not generally so.

The analysis in [21] led to the fact that among all monotone quantum Fisher information there is a smallest one which corresponds to the function $f_m(t) = (1 + t)/2$. In this case

$$F_D^{\min}(A) = \mathrm{Tr}\ AL = \mathrm{Tr}\ DL^2 \qquad \text{where} \quad DL + LD = 2A. \tag{15}$$

For the purpose of a quantum Cramér–Rao inequality the minimal quantity seems to be the best, since the inverse gives the largest lower bound. In fact, the matrix L has been used for a long time under the name of *symmetric logarithmic derivative*, see [14] and [15]. In this example the generalized covariance is

$$\varphi_D[A, B] = \tfrac{1}{2}\,\mathrm{Tr}\ D(AB + BA) \tag{16}$$

and we have

$$\mathbb{J}_D(A) = \tfrac{1}{2}(DA + AD) \qquad \text{and} \qquad \mathbb{J}_D^{-1}(A) = L = 2\int_0^\infty \mathrm{e}^{-tD} A \mathrm{e}^{-tD}\,\mathrm{d}t \tag{17}$$

for the operator $\mathbb{J}$ of the previous section.

The set of invertible $n \times n$ density matrices is a manifold of dimension $n^2 - 1$. Indeed, parametrizing these matrices by $n - 1$ real diagonal entries and $(n - 1)n/2$ upper diagonal complex entries we have $n^2 - 1$ real parameters which run over an open subset of the Euclidean space $\mathbb{R}^{n^2-1}$. Since operator monotone functions are smooth (even analytic), all the quantities γ_D in theorem 3.1 endow the manifold of density matrices with a Riemannian structure.

Fisher information appears not only as a Riemannian metric but as an information matrix as well. Let $\mathcal{M} := \{D(\theta) : \theta \in G\}$ be a smooth m-dimensional manifold of invertible density matrices. The *quantum score operators* (or logarithmic derivatives) are defined as

$$L_i(\theta) := \mathbb{J}_{D(\theta)}^{-1}\big(\partial_{\theta_i} D(\theta)\big) \qquad (1 \leqslant i \leqslant m) \tag{18}$$

and

$$I_{ij}^Q(\theta) := \mathrm{Tr}\ L_i(\theta)\mathbb{J}_{D(\theta)}\big(L_j(\theta)\big) \qquad (1 \leqslant i, j \leqslant m) \tag{19}$$

is the *quantum Fisher information matrix*.

Theorem 3.2. *Let α be a coarse graining sending density matrices on the Hilbert space $\mathcal{H}_1$ into those acting on the Hilbert space $\mathcal{H}_2$ and let $\mathcal{M} := \{D(\theta) : \theta \in G\}$ be a smooth m-dimensional manifold of invertible density matrices on $\mathcal{H}_1$. For the Fisher information matrix $I^{1Q}(\theta)$ of $\mathcal{M}$ and for Fisher information matrix $I^{2Q}(\theta)$ of $\alpha(\mathcal{M}) := \{\alpha(D(\theta)) : \theta \in G\}$ we have the monotonicity relation*

$$I^{2Q}(\theta) \leqslant I^{1Q}(\theta).$$

To prove the theorem we set $B_i(\theta) := \partial_{\theta_i} D(\theta)$. Then $\mathbb{J}_{\alpha(D(\theta))}^{-1}\alpha(B_i(\theta))$ is the score operator of $\alpha(\mathcal{M})$ and we have

$$\begin{aligned}\sum_{ij} I_{ij}^{2Q}(\theta)a_i\overline{a_j} &= \mathrm{Tr}\ \mathbb{J}_{\alpha(D(\theta))}^{-1}\alpha\Big(\sum_i a_i B_i(\theta)\Big)\alpha\Big(\sum_j \overline{a_j} B_j(\theta)\Big)\\ &\leqslant \mathrm{Tr}\ \mathbb{J}_{D(\theta))}^{-1}\Big(\sum_i a_i B_i(\theta)\Big)\Big(\sum_j \overline{a_j} B_j(\theta)\Big)\\ &= \sum_{ij} I_{ij}^{1Q}(\theta)a_i\overline{a_j}\end{aligned}$$

where (9) was used.

The monotonicity of the Fisher information matrix in some particular cases has already appeared in the literature: [20] treated the case of the Kubo–Mori inner product and [4] considered the symmetric logarithmic derivative and measurement in the role of coarse graining.

4. Garden of monotone metrics

All the monotone quantum Fisher information quantities in the range of the previous theorem are smoothly dependent on the footpoint density D and hence they endow the state space with a *Riemannian structure*. In particular, the Riemannian geometry of the minimal Fisher information was the subject of the paper [6].

It is instructive to consider the state space of a 2-level quantum system in detail. Dealing with 2×2 density matrices, we conveniently use the so-called Stokes parametrization

$$D_x = \tfrac{1}{2}(I + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3) \equiv \tfrac{1}{2}(I + x \cdot \sigma) \tag{20}$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices and $(x_1, x_2, x_3) \in \mathbb{R}^3$ with $x_1^2 + x_2^2 + x_3^2 \leqslant 1$. A monotone Fisher information on $\mathcal{M}_2$ is rotation invariant in the sense that it depends only on $r = \sqrt{x^2 + y^2 + z^2}$ and splits into radial and tangential components as follows

$$\mathrm{d}s^2 = \frac{1}{1-r^2}\,\mathrm{d}r^2 + \frac{1}{1+r}\,g\left(\frac{1-r}{1+r}\right)\mathrm{d}n^2 \qquad \text{where} \quad g(t) = \frac{1}{f(t)}. \tag{21}$$

The radial component is independent of the function f. (This fact is again a reminder of the Cencov uniqueness theorem.) The limit of the tangential component exists in (21) when $r \to 1$ provided that $f(0) \neq 0$. In this way the standard Fubini–Study metric is obtained on the set of pure states, up to a constant factor. (In the case of larger density matrices, pure states form a small part of the topological boundary of the invertible density matrices. Hence, in order to speak about the extension of a Riemannian metric on invertible densities to pure states, a rigorous meaning of the extension should be given. This is the subject of the paper [28], see also [25].) Besides minimality the radial extension yields another characterization of the minimal quantum Fisher information, see [25].

Theorem 4.1. *Among the monotone quantum Fisher information the minimal one (given by (15)) is characterized by the properties that it is classically Fisher-adjusted (in the sense of (12)) and its radial limit is the Fubini–Study metric on pure states.*

We note that in the minimal case $f_m(t) = (t+1)/2$ we have a constant tangential component in (21):

$$\mathrm{d}s^2 = \frac{1}{1-r^2}\,\mathrm{d}r^2 + \mathrm{d}n^2. \tag{22}$$

The metric (15) is widely accepted in the role of quantum Fisher information, see [2]. However, some other operator monotone functions may also have importance. Let us first view the other extreme. According to [21] there is a largest metric among all the monotone quantum Fisher information and this corresponds to the function $f_M(t) = 2t/(1+t)$. In this case

$$\mathbb{J}_D^{-1}(A) = \tfrac{1}{2}(D^{-1}A + AD^{-1}) \qquad \text{and} \qquad F_D^{\max}(A) = \mathrm{Tr}\ D^{-1}A^2. \tag{23}$$

The maximal metric cannot be extended to pure states.

It can be proved that the function

$$f_\beta(t) = \beta(1-\beta)\frac{(x-1)^2}{(x^\beta - 1)(x^{1-\beta} - 1)} \tag{24}$$

is operator monotone. This was done for the case $0 < \beta < 1$ in [22] and the case $-1 < \beta < 0$ was treated in [13]. (The operator monotonicity also follows from (35).) We denote by F^β the corresponding Fisher information metric. When $A = \mathrm{i}[D, B]$ is orthogonal to the commutator of the footpoint D in the tangent space, we have

$$F_D^\beta(A) = \frac{1}{2\beta(1-\beta)}\,\mathrm{Tr}\,\big([D^\beta, B][D^{1-\beta}, B]\big). \tag{25}$$

Apart from a constant factor this expression is the skew information proposed by Wigner and Yanase some time ago [29]. In the limiting cases $\beta \to 0$ or 1 we have

$$f_0(x) = \frac{1-x}{\log x}$$

and the corresponding metric

$$K_D(A, B) := \int_0^\infty \mathrm{Tr}\ A(D+t)^{-1}B(D+t)^{-1}\,\mathrm{d}t \tag{26}$$

is named after Kubo, Mori, Bogoliubov, etc. The Kubo–Mori inner product plays a role in quantum statistical mechanics (see [8], for example). In this case

$$\mathbb{J}^{-1}(B) = \int_0^\infty (D+t)^{-1}B(D+t)^{-1}\,\mathrm{d}t \qquad \text{and} \qquad \mathbb{J}(A) = \int_0^1 D^t A D^{1-t}\,\mathrm{d}t. \tag{27}$$

Therefore the corresponding generalized variance is

$$\varphi_D(A, B) = \int_0^1 \mathrm{Tr}\ AD^tBD^{1-t}\,\mathrm{d}t. \tag{28}$$

Beyond the affine parametrization of the set of density matrices, the exponential parametrization is another possibility: any density matrix is written in a unique way in the form $\mathrm{e}^H/\,\mathrm{Tr}\ \mathrm{e}^H$, where H is a self-adjoint traceless matrix. In the affine parametrization the integral (26) gives the metric and (28) is the corresponding variance. If we substitute for the exponential parametrization, the role of the two formulae is interchanged: integral (26) gives the variance and (28) is the metric. (The reason for this is the fact that the change of the coordinates is described by $\mathbb{J}$ from (27).) The affine and exponential parametrization is the subject of the paper [12] and the characterization of the Kubo–Mori metric in [11] is probably another form of the duality observed between (26) and (28).

5. The Cramér–Rao inequalities revisited

Let $\mathcal{M} := \{D_\theta : \theta \in G\}$ be a smooth m-dimensional manifold and assume that a collection $A = (A_1, \ldots, A_m)$ of self-adjoint matrices is used to estimate the true value of θ.

Given an operator $\mathbb{J}$ we have the corresponding generalized variance φ_θ for every θ and the generalized covariance matrix of the estimator A is a positive definite matrix, defined by $\varphi_\theta[A]_{ij} = \varphi_\theta[A_i, A_j]$. The *bias* of the estimator is

$$\begin{aligned} b(\theta) &= \big(b_1(\theta), b_2(\theta), \ldots, b_m(\theta)\big) \\ &:= \big(\mathrm{Tr}\ D_\theta(A_1-\theta_1), \mathrm{Tr}\ D_\theta(A_2-\theta_2), \ldots, \mathrm{Tr}\ D_\theta(A_m-\theta_m)\big). \end{aligned}$$

For an *unbiased estimator* we have $b(\theta) = 0$. From the bias vector we form a bias matrix

$$B_{ij}(\theta) := \partial_{\theta_i} b_j(\theta).$$

For a *locally unbiased estimator* at θ_0, we have $B(\theta_0) = 0$.

The relation

$$\partial_{\theta_i}\ \mathrm{Tr}\ D_\theta H = \varphi_\theta[L_i(\theta), H]$$

determines the logarithmic derivatives $L_i(\theta)$. The Fisher information matrix is

$$I_{ij}(\theta) := \varphi_\theta[L_i(\theta), L_j(\theta)].$$

Theorem 5.1. *Let $A = (A_1, \ldots, A_m)$ be an estimator of θ. Then for the previously defined quantities the inequality*

$$\varphi_\theta[A] \geqslant \big(I + B(\theta)\big)J(\theta)^{-1}\big(I + B(\theta)^*\big)$$

holds in the sense of the order on positive semidefinite matrices.

The proof is rather simple if we use the block matrix method. Let X and Y be $m \times m$ matrices with $n \times n$ entries and assume that all entries of Y are constant multiples of the unit matrix. (A_i and L_i are $n \times n$ matrices.) If α is a completely positive mapping on $n \times n$ matrices, then $\tilde{\alpha} := \mathrm{Diag}\,(\alpha, \ldots, \alpha)$ is a positive mapping on block matrices and $\tilde{\alpha}(YX) = Y\tilde{\alpha}(X)$. This implies that $\mathrm{Tr}\ X\alpha(X^*)Y \geqslant 0$ when Y is positive. Therefore the $m \times m$ ordinary matrix M which has ij entry

$$\mathrm{Tr}\,(X\tilde{\alpha}(X^*))_{ij}$$

is positive. In the sequel we restrict ourselves to $m = 2$ for the sake of simplicity and apply the above fact to the case

$$X = \begin{bmatrix} A_1 & 0 & 0 & 0 \\ A_2 & 0 & 0 & 0 \\ L_1(\theta) & 0 & 0 & 0 \\ L_2(\theta) & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad \alpha = \mathbb{J}_{D(\theta)}.$$

Then we have

$$M = \begin{bmatrix} \mathrm{Tr}\ A_1\mathbb{J}_D(A_1) & \mathrm{Tr}\ A_1\mathbb{J}_D(A_2) & \mathrm{Tr}\ A_1\mathbb{J}_D(L_1) & \mathrm{Tr}\ A_1\mathbb{J}_D(L_2) \\ \mathrm{Tr}\ A_2\mathbb{J}_D(A_1) & \mathrm{Tr}\ A_2\mathbb{J}_D(A_2) & \mathrm{Tr}\ A_2\mathbb{J}_D(L_1) & \mathrm{Tr}\ A_2\mathbb{J}_D(L_2) \\ \mathrm{Tr}\ L_1\mathbb{J}_D(A_1) & \mathrm{Tr}\ L_1\mathbb{J}_D(A_2) & \mathrm{Tr}\ L_1\mathbb{J}_D(L_1) & \mathrm{Tr}\ L_1\mathbb{J}_D(L_2) \\ \mathrm{Tr}\ L_2\mathbb{J}_D(A_1) & \mathrm{Tr}\ L_2\mathbb{J}_D(A_2) & \mathrm{Tr}\ L_2\mathbb{J}_D(L_1) & \mathrm{Tr}\ L_2\mathbb{J}_D(L_2) \end{bmatrix} \geqslant 0.$$

.

Now we rewrite the matrix M in terms of the matrices involved in our Cramér–Rao inequality. The 2×2 block M_{11} is the generalized covariance, M_{22} is the Fisher information matrix and M_{12} is easily expressed as $I + B$. We obtain

$$M = \begin{bmatrix} \varphi_\theta[A_1, A_1] & \varphi_\theta[A_1, A_2] & 1 + B_{11}(\theta) & B_{12}(\theta) \\ \varphi_\theta[A_2, A_1] & \varphi_\theta[A_2, A_2] & B_{21}(\theta) & 1 + B_{22}(\theta) \\ 1 + B_{11}(\theta) & B_{21}(\theta) & \varphi_\theta[L_1, L_1] & \varphi_\theta[L_1, L_2] \\ B_{12}(\theta) & 1 + B_{22}(\theta) & \varphi_\theta[L_2, L_1] & \varphi_\theta[L_2, L_2] \end{bmatrix} \geqslant 0.$$

Since the positivity of a block matrix

$$M = \begin{bmatrix} M_1 & C \\ C^* & M_2 \end{bmatrix} = \begin{bmatrix} \varphi_D[A] & I + B \\ I + B^* & J(\theta) \end{bmatrix}$$

implies $M_1 \geqslant CM_2^{-1}C^*$ we have exactly the statement of our Cramér–Rao inequality.

6. Statistical distinguishability and uncertainty

Assume that a manifold $\mathcal{M} := \{D_\theta : \theta \in G\}$ of density matrices is given together with a statistically relevant Riemannian metric γ_d. We do not give a formal definition of such a metric. What we have in mind is the property that given two points on the manifold their geodesic distance is interpreted as the statistical distinguishability of the two density matrices in some statistical procedure.

Let $D_0 \in \mathcal{M}$ be a point on our statistical manifold. The geodesic ball

$$B_\varepsilon(D_0) := \{D \in \mathcal{M} : d(D_0, D) < \varepsilon\}$$

contains all density matrices which can be distinguished by an effort smaller than ε from the fixed density D_0. The size of the inference region $B_\varepsilon(D_0)$ measures the statistical uncertainty at the density D_0. Following *Jeffrey's rule* the size is the volume measure determined by the statistical (or information) metric. More precisely, it is better to consider the asymptotics of the volume of $B_\varepsilon(D_0)$ as $\varepsilon \to 0$. According to differential geometry

$$\mathrm{Vol}\big(B_\varepsilon(D_0)\big) = C_n\varepsilon^n - \frac{C_n}{6(n+2)}\mathrm{Scal}\,(D_0)\varepsilon^{n+2} + \mathrm{o}(\varepsilon^{n+2}) \qquad (29)$$

where n is the dimension of our manifold, C_n is a constant (equal to the volume of the unit ball in the Euclidean n-space) and Scal means the scalar curvature, see theorem 3.98 in [10]. In this way, the scalar curvature of a statistically relevant Riemannian metric might be interpreted as the *average statistical uncertainty* of the density matrix (in the given statistical manifold). This interpretation becomes particularly interesting for the full state space endowed by the Kubo–Mori inner product as a statistically relevant Riemannian metric.

Let $\mathcal{M}$ be the manifold of all invertible $n \times n$ density matrices. The Kubo–Mori (or Bogoliubov) inner product is given by

$$\gamma_D(A, B) = \mathrm{Tr}\,(\partial_A D)(\partial_B \log D). \tag{30}$$

In particular, in the affine parametrization we have

$$\gamma_D(A, B) = \int_0^\infty \mathrm{Tr}\ A(D+t)^{-1}B(D+t)^{-1} \tag{31}$$

see [20]. On the basis of numerical evidence it was *conjectured* in [20] that the scalar curvature which is a statistical uncertainty is monotone in the following sense. For any coarse graining α the scalar curvature at a density D is smaller than at $\alpha(D)$. The average statistical uncertainty is increasing under coarse graining. Up to now this conjecture has not been proven mathematically. Another form of the conjecture is the statement that along a curve of Gibbs states

$$\frac{\mathrm{e}^{-\beta H}}{\mathrm{Tr}\ \mathrm{e}^{-\beta H}}$$

the scalar curvature changes monotonically with the inverse temperature $\beta \geqslant 0$, that is, *the scalar curvature is a monotone decreasing function of* β.

7. Relative entropy as a contrast function

Let D_θ be a smooth manifold of density matrices. The following construction is motivated by classical statistics. Suppose that a non-negative functional $d(D_1, D_2)$ of two variables is given on the density matrices. In many cases one can obtain a Riemannian metric by differentiation

$$g_{ij}(\theta) = \frac{\partial^2}{\partial\theta_i\,\partial\theta'_j} d(D_\theta, D_{\theta'})\Big|_{\theta=\theta'}.$$

To be more precise the non-negative smooth functional $d(\,\cdot\,,\,\cdot\,)$ is called a contrast functional if $d(D_1, D_2) = 0$ implies $D_1 = D_2$. (For the role of contrast functionals in classical estimation, see [7].) We note that a contrast functional is a particular example of yokes, cf [3].

Following the work of Csiszár in classical information theory, Petz introduced a family of information quantities parametrized by a function $F : \mathbb{R}^+ \to \mathbb{R}$

$$S_F(D_1, D_2) = \mathrm{Tr}\,\left(D_1^{1/2} F(\Delta_{D_2,D_1}) D_1^{1/2}\right) \tag{32}$$

see [19], or [18, p 113]. Here $\Delta_{D_2,D_1} := L_{D_2} R_{D_1}^{-1}$ is the relative modular operator of the two densities. When F is operator convex, this quasi-entropy possesses good properties, for example, it is a contrast functional in the above sense if F is not linear. In particular for

$$F(t) = \frac{4}{1-\alpha^2}\left(1 - t^{(1+\alpha)/2}\right)$$

we have

$$S_\alpha(D_1, D_2) = \frac{4}{1-\alpha^2}\,\mathrm{Tr}\,\left(I - D_2^{\frac{1+\alpha}{2}} D_1^{-\frac{1+\alpha}{2}}\right) D_1. \tag{33}$$

By differentiating we obtain

$$\frac{\partial^2}{\partial t \partial u} S_\alpha(D + tA, D + uB)\Big|_{t=u=0} = K_D^\alpha(A, B) \tag{34}$$

which is related to (25) as

$$F_D^\beta(A) = K_D^\alpha(A, A) \qquad \text{and} \qquad \beta = (1-\alpha)/2.$$

Ruskai and Lesniewski discovered that all monotone Fisher information is obtained from a quasi-entropy as the contrast functional [17]. The relation of the function F in (32) to the function f in theorem 3.1 is

$$\frac{1}{f(t)} = \frac{F(t) + tF(t^{-1})}{(t-1)^2}. \tag{35}$$

Acknowledgments

This work was partially supported by EPSRC grant GR/R53234/01 and by the Hungarian OTKA T032662.

References

[1] Amari S 1985 *Differential-Geometrical Methods in Statistics (Lecture Notes in Stat. vol 28)* (Berlin: Springer)
[2] Barndorff-Nielsen O E and Gill R D 2000 Fisher information in quantum statistics *J. Phys. A: Math. Gen.* **33** 4481–90
[3] Barndorff-Nielsen O E and Jupp P E 1997 Yokes and symplectic structures *J. Stat. Plan. Inference* **63** 133–46
[4] Braunstein S L and Caves C M 1994 Statistical distance and the geometry of quantum states *Phys. Rev. Lett.* **72** 3439–43
[5] Cencov N N 1982 *Statistical Decision Rules and Optimal Inferences (Translation of Math. Monog. vol 53)* (Providence, RI: American Mathematical Society)
[6] Dittmann J 1993 On the Riemannian geometry of finite-dimensional state space *Semin. Sophus Lie* **3** 73–87
[7] Eguchi S 1983 Second order efficiency of minimum contrast estimation in a curved exponential family *Ann. Stat.* **11** 793–803
[8] Fick E and Sauermann G 1990 *The Quantum Statistics of Dynamic Processes* (Berlin: Springer)
[9] Fujiwara A and Nagaoka H 1995 Quantum Fisher metric and estimation for pure state models *Phys. Lett.* A **201** 119–24
[10] Gallot S, Hulin D and Lafontaine J 1993 *Riemannian Geometry* (Berlin: Springer)
[11] Grasselli M R and Streater R F 2001 On the uniqueness of the Chentsov metric in quantum information geometry *Infinite-Dimensional Anal. Quantum Prob.* **4** 173–82
[12] Hasegawa H 1997 Exponential and mixture families in quantum statistics: dual structure and unbiased parameter estimation *Rep. Math. Phys.* **39** 49–68
[13] Hasegawa H and Petz D 1997 *Non-commutative Extension of Information Geometry II Quantum Communication, Computing and Measurement* ed Hirota *et al* (New York: Plenum)
[14] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[15] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[16] Kass R E 1989 The geometry of asymptotic inference *Stat. Sci.* **4** 188–234
[17] Lesniewski A and Ruskai M B 1999 Monotone Riemannian metrics and relative entropy on noncommutative probability spaces *J. Math. Phys.* **40** 5702–24
[18] Ohya M and Petz D 1993 *Quantum Entropy and its Use* (Heidelberg: Springer)
[19] Petz D 1986 Quasi-entropies for finite quantum systems *Rep. Math. Phys.* **23** 57–65
[20] Petz D 1994 Geometry of canonical correlation on the state space of a quantum system *J. Math. Phys.* **35** 780–95
[21] Petz D 1996 Monotone metrics on matrix spaces *Linear Algebr. Appl.* **244** 81–96
[22] Petz D and Hasegawa H 1996 On the Riemannian metric of α-entropies of density matrices *Lett. Math. Phys.* **38** 221–5
[23] Petz D and Jenčova A 1999 On quantum Fisher information *J. Electr. Eng.* **50** 78–81
[24] Petz D and Sudár Cs 1996 Geometries of quantum states *J. Math. Phys.* **37** 2662–73

[25] Petz D and Sudár C 1999 Extending the Fisher metric to density matrices *Geometry of Present Days Science* ed O E Barndorff-Nielsen and E B Vendel Jensen (Singapore: World Scientific) pp 21–34
[26] Slater P 1998 Comparative noninformativities of quantum priors based on monotone metrics *Phys. Lett.* A **247** 1–8
[27] Streater R F 2000 Classical and quantum info-manifolds *Preprint* math-ph/0002050
[28] Sudár C 1996 Radial extension of monotone Riemannian metrics on density matrices *Publ. Math. Debrecen* **49** 243–50
[29] Wigner E P and Yanase M M 1963 Information content of distributions *Proc. Natl Acad. Sci. USA* **49** 910–18