On the philosophy of Cramér-Rao-Bhattacharya Inequalities in Quantum Statistics

To my revered guru C. R. Rao for revealing the mystries of Chance

K. R. Parthasarathy

Indian Statistical Institute

Delhi Centre,

7, S. J. S. Sansanwal Marg,

New Delhi - 110 016, India

e-mail: krp@isid.ac.in

Summary   To any parametric family of states of a finite level quantum system we associate a space of Fisher maps and introduce the natural notions of Cramér-Rao-Bhattacharya tensor and Fisher information form. This leads us to an abstract Cramér-Rao-Bhattacharya lower bound for the covariance matrix of any finite number of unbiased estimators of parameteric functions. A number of illustrative examples is included. Modulo technical assumptions of various kinds our methods can be applied to infinite level quantum systems as well as parametric families of classical probability distributions on Borel spaces.

Key words: Finite level quantum system, uncertainty principle, generalized measurement, Covariance matrix of unbiased estimators, Fisher map, Fisher information form, Cramér-Rao-Bhattacharya tensor, Cramér-Rao-Bhattacharya bound.

AMS Subject classification index: 81C20, 94A15.

1. Introduction

The evolution of modern scientific thought is strewn with several examples expressing the following sentiment: in any effort to accomplish a task there can be a certain limit to the efficiency of its performance. In the present context we bring to attention three such famous examples which are based on the combination of a deep conceptual approach and simple mathematical arguments. Finally, we shall focus on one of them, namely, limits to the efficiency of estimating an unknown parameter involved in a family of states of a finite level quantum system.

Our first example is the celebrated uncertainty principle of Heisenberg [8] in quantum mechanics. For an interesting historical account of this great discovery in the philosophy of science we refer the reader to the essay by Jagdish Mehra

1

in [13]. If q and p denote the position and momentum operators of a quantum mechanical particle executing motion on the real line $\mathbb{R}$ they obey the commutation relation $qp - pq = i\hbar$ where $\hbar = h/2\pi$, h being the Planck's constant, and this implies the following inequality. If ψ is the absolutely square integrable wave function describing the state of the system and $\mathrm{Var}(X|\psi)$ denotes the variance of the observable in the state ψ then

$$\mathrm{Var}(q|\psi)\mathrm{Var}(p|\psi) \geqslant \hbar^2/4. \tag{1.1}$$

In particular, if the variance of p in the state ψ is σ^2 then

$$\mathrm{Var}(q|\psi) \geqslant \hbar^2/4\sigma^2.$$

In other words, this sets a limit to the accuracy with which the position q can be measured in the state ψ. Such limits to accuracy hold for any 'conjugate pair' of observables in quantum theory.

Our second example is the famous Cramér-Rao inequality [4], [18] in the theory of estimation of statistical parameters. For an amusing and insightful account of the route by which this fundamental discovery was made and how it came to be recognized in the history of statistical science we refer to [19]. Suppose $\{p(\omega, \theta)\}$ is a parametric family of probability density functions with respect to a σ-finite measure in a Borel space $(\Omega, \mathcal{F})$, θ being a real parameter varying in an open interval (a, b). Assume that the function

$$I(\theta) = \int_\Omega \left(\frac{\partial}{\partial \theta} \log p(\omega, \theta) \right)^2 p(\omega, \theta)\mu(d\omega) \tag{1.2}$$

is well-defined for all θ in (a, b). On the basis of a sample point ω obtained from experiment evaluate a function $T(\omega)$ as an estimate of the parameter θ. The function $T(\cdot)$ on Ω is called an *unbiased estimator* of θ if

$$\int_\Omega T(\omega)p(\omega, \theta)\mu(d\omega) = \theta \quad \forall \quad \theta \in (a, b)$$

and, in such a case, its variance, denoted by $V(T|\theta)$ is defined by

$$V(T|\theta) = \int (T(\omega) - \theta)^2 \ p(\omega, \theta)\mu(d\omega).$$

Indeed, $V(T|\theta)$ is a measure of the error involved in estimating θ by $T(\omega)$. The Cramér-Rao inequality in its simplest form says that

$$V(T|\theta) \geqslant I(\theta)^{-1} \tag{1.3}$$

where $I(\theta)$ is given by (1.2) and called the 'Fisher information' at θ. Thus (1.3) sets a limit to the accuracy of estimating the unknown parameter θ from experimental observation.

It is a remarkable fact that a special case of (1.3) implies the Heisenberg uncertainty principle (1.1) and much more in emphasizing the profundity of Fisher information. Indeed, let $\psi \in L^2(\mathbb{R})$ be a wave function so that $\|\psi\| = 1$. By changing ψ to a new wave function $e^{i\alpha x}\psi(x)$, $\alpha \in \mathbb{R}$, if necessary, we may assume, without loss of generality that the momentum operator p satisfies the condition $\langle \psi|p|\psi \rangle = 0$ and $\langle \psi|q|\psi \rangle = m$, a real scalar. By Born's interpretation $f = |\psi|^2$ is

the probability density function of the position observable q in the state ψ. Introducing the parametric family $\{f(x-\theta), \theta \in \mathbb{R}\}$ of probability densities we see that its Fisher information $I(\theta)$ is given by

$$
\begin{aligned}
I(\theta) &= \int_{\mathbb{R}} \left(\frac{f'(x-\theta)}{f(x-\theta)} \right)^2 f(x-\theta) \, dx \\
&= 4 \int_{\mathbb{R}} \left(\mathrm{Re} \, \frac{\psi'}{\psi}(x) \right)^2 |\psi(x)|^2 \, dx
\end{aligned}
\tag{1.4}
$$

and therefore independent of θ. By Cramér-Rao inequality we have

$$
\begin{aligned}
\mathrm{Var}(q|\psi) &= \int (x-m)^2 \, f(x) \, dx \\
&\geqslant \frac{1}{I(m)}.
\end{aligned}
\tag{1.5}
$$

On the other hand

$$
\mathrm{Var}(p|\psi) = \int_{\mathbb{R}} x^2 \, |(F\psi)(x)|^2 \, dx
$$

where F is the unitary Fourier transform in $L^2(\mathbb{R})$. Thus by (1.4) we have

$$
\begin{aligned}
\mathrm{Var}(p|\psi) &= \langle \psi | F^\dagger q^2 F | \psi \rangle \\
&= \|p\psi\|^2 \\
&= \hbar^2 \int |\psi'(x)|^2 \, dx \\
&= \hbar^2 \int \left| \frac{\psi'}{\psi}(x) \right|^2 |\psi(x)|^2 \, dx \\
&\geqslant \hbar^2 \int \left| \left(\mathrm{Re} \, \frac{\psi'}{\psi} \right)(x) \right|^2 |\psi(x)|^2 \, dx \\
&= \frac{\hbar^2}{4} \, I(m)
\end{aligned}
\tag{1.6}
$$

which together with (1.5) implies (1.1). The more powerful inequality (1.6) and its natural generalization for covariance matrices in $L^2(\mathbb{R}^n)$ are known together as Stam's uncertainty principle. For more information along these lines and a rich survey of information inequalities we refer to the paper [5] by A. Dembo, T. M. Cover and J. A. Thomas.

Our last illustrious example is of a different genre but again connected with the notion of information. It is Shannon's noisy coding theorem [20] which sets a limit to the ability of communication through an information channel in the presence of noise. Again we present the simplest version of this strikingly beautiful result in order to highlight the philosophical aspect and refer to [17] for more general versions. Consider an information channel whose input and output alphabets are same and equal to the binary alphabet $\{0, 1\}$ which is also a field of two elements with the operations of addition and multiplication modulo 2. If an input letter x from this alphabet is transmitted through the channel assume that the output

letter is x or $x + 1$ with probability $1 - p$ or p so that the probability of error due to noise in transmission is p. Such a channel is said to be *binary* and *symmetric*. Assume that the transmission of a sequence $x_1, x_2, \ldots, x_n$ of n letters through this binary symmetric channel yields the output sequence $y_1, y_2, \ldots, y_n$ where $y_1 - x_1$, $y_2 - x_2, \ldots, y_n - x_n$ are independently and identically distributed Bernoulli random variables, each assuming the values 0 and 1 with probability $q = 1 - p$ and p respectively. Such a channel is called a *memoryless* binary symmetric channel.

Denote the alphabet by $\mathbb{F}_2$. By a *code* of *size m*, *length n* and *error probability* not exceeding ε, where $0 < \varepsilon < 1$, we mean m pairs $(u_j, E_j), 1 \leqslant j \leqslant m$, $u_j \in \mathbb{F}_2^n$, $E_j \subset \mathbb{F}_2^n$, $E_1, E_2, \ldots, E_m$ are pairwise disjoint, satisfying the inequalities

$$\mathbb{P}\left(\text{output sequence } \in E_j \middle| \text{input sequence } = u_j\right) > 1 - \varepsilon \quad \forall j.$$

Denote by $N(n, p, \varepsilon)$ the maximum possible size for codes of length n with error probability not exceeding ε. Then

$$\lim_{n \to \infty} \frac{1}{n} \log_2 N(n, p, \varepsilon) = 1 + p \log_2 p + q \log_2 q \quad \forall \, 0 < \varepsilon < 1, \, 0 < p < 1/2. \quad (1.7)$$

If we write $H(p) = -p \log_2 p - q \log_2 q$ and call it the Shannon entropy of the Bernoulli random variable with probability of success (error) p then (1.7) has the interpretation that for large n, among the 2^n possible input sequences of length n roughly $2^{n(1-H(p))}$ sequences could be transmitted with error probability $< \varepsilon$ and not more. For this reason the expression on the right hand side of (1.7) is called the Shannon capacity of the binary symmetric channel with error probability p. A corresponding generalization for memoryless and stationary quantum channels describing their 'capacity' to transmit classical alphabetic messages exists. For a leisurely and self-contained exposition of such coding theorems see [17]. The notion of entropy that arises in the brief discussion above can be introduced for a large class of density functions and this, in turn, leads to some remarkable connections with Fisher information and many powerful information theoretic inequalities. Once again we refer to the very rich survey article [5].

All the three examples described above have been generalized in several ways, connections between them and relations with other branches of science and engineering have emerged and an enormous amount of literature has grown around them. The last example has given birth to the subject of quantum information theory and coding theorems for quantum channels [14], [17]. The present essay is devoted to the second example but in the context of parametric families of states of finite level quantum systems. Starting from the books of Helström [9], Holevo [10], and Hayashi [7] there is quite some literature on the Cramér Rao bounds for quantum systems. By confining ourselves to finite level systems we avoid the technical difficulties of dealing with unbounded operators and their varying domains but we gain conceptual and algebraic clarity.

In Section 2 we give a brief account of the quantum probability of finite level quantum systems in a complex finite dimensional Hilbert space including the notions of events, observables, states, generalized measurements and composite systems in the language of tensor products of Hilbert spaces. Heisenberg's uncertainty principle and an entropic uncertainty principle are briefly described. The notions

of parametric families of states and unbiased estimators of parametric functions along with their variances and covariances are introduced.

Section 3 contains the key notions, namely, Fisher maps, the Fisher information form and the Cramér-Rao-Bhattacharya (CRB) tensor with respect to a parametric family of states of a finite level quantum system. The Cramér-Rao-Bhattacharya (CRB) bound is finally expressed in terms of the CRB tensor and the Fisher information form. Several illustrative examples are given.

In the last section we show how, by using a dilation theorem of Naimark, one can obtain a CRB bound for the covariance matrix of unbiased estimators of parametric functions based on generalized measurements.

2. Preliminaries in the quantum probability and statistics of finite level systems

A finite level quantum system is described by 'states' in a finite dimensional complex Hilbert space. We choose and fix such a Hilbert space $\mathcal{H}$ with scalar product $\langle u|v \rangle$ which is linear in the variable v and antilinear in u. A typical example obtains when $\mathcal{H}$ is the n-dimensional complex vector space $\mathbb{C}^n$ of column vectors and its dual is the space of all row vectors. In this case the scalar product is expressed as

$$\langle u|v \rangle = \sum_{i=1}^{n} \bar{a}_i b_i$$

where

$$u = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, v = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad a_i, b_i \in \mathbb{C} \; \forall \; i.$$

Elements of $\mathcal{H}$ are called *ket* vectors, a typical element in $\mathcal{H}$ being denoted by $|v\rangle$ whereas any element in the dual of $\mathcal{H}$ is called a *bra* vector and a typical bra vector is denoted by $\langle u|$. The linear functional $\langle u|$ evaluated at a ket vector $|v\rangle$ is the scalar product $\langle u|v \rangle$. If A is an operator in $\mathcal{H}$ it is customary to write

$$\langle u|Av \rangle = \langle u|A|v \rangle.$$

The adjoint of A is denoted by $A^\dagger$ so that

$$\langle u|A|v \rangle = \langle A^\dagger u|v \rangle = \langle u|Av \rangle.$$

In such a notation $|u\rangle\langle v|$ denotes the operator satisfying

$$(|u\rangle\langle v|) \; |w\rangle = \langle v|w\rangle|u\rangle \quad \forall \quad |u\rangle, |v\rangle, |w\rangle \text{ in } \mathcal{H}.$$

The trace of an operator A in $\mathcal{H}$ is denoted by Tr A. In particular Tr $|u\rangle\langle v| = \langle v|u\rangle$. Note that $|u\rangle\langle v|$ is a rank one operator when $|u\rangle \neq 0$, $|v\rangle \neq 0$, and

$$(|u_1\rangle\langle v_1|) (|u_2\rangle\langle v_2|) \cdots (|u_k\rangle\langle v_k|) = c|u_1\rangle\langle v_k|$$

where $c = \langle v_1|u_2\rangle\langle v_2|u_3\rangle \cdots \langle v_{k-1}|u_k\rangle$.

Denote by $\mathcal{B}(\mathcal{H})$, $\mathcal{P}(\mathcal{H})$, $\mathcal{O}(\mathcal{H})$, $\mathcal{S}(\mathcal{H})$ respectively the $*$- algebra of all operators on $\mathcal{H}$ with its usual (strong) topology, the orthomodular lattice of all orthogonal

projection operators on $\mathcal{H}$, the real linear space of all hermitian operators in $\mathcal{H}$ and the compact convex set of all nonnegative definite operators of unit trace. We have $\mathcal{P}(\mathcal{H}) \subset \mathcal{O}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ and $\mathcal{S}(\mathcal{H}) \subset \mathcal{O}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$. If $A, B \in \mathcal{O}(\mathcal{H})$ we say that $A \leqslant B$ if $B - A$ is nonnegative definite. Then $\mathcal{O}(\mathcal{H})$ is a partially ordered real linear space. A nonnegative definite hermitian operator is simply called a positive operator.

The zero and identity operators are denoted respectively by O amd I. Often, I is denoted by 1. For any scalar λ the operator λI is also denoted by λ. Thus, for $A \in \mathcal{B}(\mathcal{H})$, $\lambda \in \mathbb{C}$, $A - \lambda$ stands for the operator $A - \lambda I$. For any $E \in \mathcal{P}(\mathcal{H})$, $0 \leqslant E \leqslant 1$ and $(1 - E) \in \mathcal{P}(\mathcal{H})$. By a projection we shall always mean an orthogonal projection operator i.e., an element of $\mathcal{P}(\mathcal{H})$. If $E_1, E_2 \in \mathcal{P}(\mathcal{H})$ and $E_1 \leqslant E_2$ then $(E_2 - E_1) \in \mathcal{P}(\mathcal{H})$. When a quantum system is described by $\mathcal{H}$ we say that the elements of $\mathcal{P}(\mathcal{H})$ are the *events* concerning the system, 0 is the *null* event and 1 is the *certain* event. If $E_1, E_2 \in \mathcal{P}(\mathcal{H})$ and $E_1 \leqslant E_2$ we say that the event E_1 *implies* the event E_2. If $E \in \mathcal{P}(\mathcal{H})$ then $1 - E$ is the event 'not E'. If $E_1, E_2 \in \mathcal{P}(\mathcal{H})$ their maximum $E_1 \vee E_2$ and minimum $E_1 \wedge E_2$ are respectively interpreted as 'E_1 or E_2' and 'E_1 and E_2'. If $E_1 E_2 = 0$ then $E_1 \vee E_2 = E_1 + E_2$. If E_1 and E_2 commute then $E_1 \wedge E_2 = E_1 E_2$. The first basic difference between classical probability and quantum probability theory arises from the fact that for three events E_i in $\mathcal{P}(\mathcal{H}), i = 1, 2, 3$ one may not have $E_1 \wedge (E_2 \vee E_3) = (E_1 \wedge E_2) \vee (E_1 \wedge E_3)$. Whenever the E_i's commute with each other the operations $\wedge$ and $\vee$ distribute with each other.

Any hermitian opearator X in $\mathcal{H}$ is called a real-valued or simply an *observable* about the system described by $\mathcal{H}$. Thus $\mathcal{O}(\mathcal{H})$ is the real linear space of all real-valued observables. If $X, Y \in \mathcal{O}(\mathcal{H})$ and $XY = YX$ then XY is also an element of $\mathcal{O}(\mathcal{H})$. If $X \in \mathcal{O}(\mathcal{H})$ and $\sigma(X)$ denotes the set of all its eigenvalues then, by the spectral theorem, X admits a unique spectral resolution or representation

$$X = \sum_{\lambda \in \sigma(X)} \lambda\, E_\lambda \tag{2.1}$$

where $\sigma(X) \subset \mathbb{R}$ is a finite set of cardinality not exceeding the dimension of $\mathcal{H}$, $0 \neq E_\lambda \in \mathcal{P}(\mathcal{H}) \ \forall\ \lambda \in \sigma(X)$ and

$$\sum_{\lambda \in \sigma(X)} E_\lambda \ = \ I, \tag{2.2}$$

$$E_\lambda E_{\lambda'} \ = \ \delta_{\lambda\lambda'} E_\lambda \quad \forall \quad \lambda, \lambda' \in \sigma(X). \tag{2.3}$$

This, at once, suggests the interpretation that the eigenprojection E_λ associated with the eigenvalue λ in (2.1) is the event that the observable X takes the value λ and $\sigma(X)$ is the set of all values that X can take. If $\varphi : \sigma(X) \to \mathbb{R}$ or $\mathbb{C}$ is a real or complex-valued function then

$$\varphi(X) = \sum_{\lambda \in \sigma(X)} \varphi(\lambda)\, E_\lambda \tag{2.4}$$

is the real or complex-valued observable which is the function φ of X.

Any element $\rho \in \mathcal{S}(\mathcal{H})$ is called a *state* of the quantum system described by X. Such a state ρ is also called a *density operator.* Clearly, ρ itself becomes an observable. If $E \in \mathcal{P}(\mathcal{H})$ is an event and ρ is a state then $\operatorname{Tr} \rho E$ is a quantity in the unit interval $[0, 1]$ called the *probability* of the event E in the state ρ. If E_1, E_2 are two events satisfying the relation $E_1 E_2 = 0$ then $E_1 + E_2$ is also an event and $\operatorname{Tr} \rho(E_1 + E_2) = \operatorname{Tr} \rho E_1 + \operatorname{Tr} \rho E_2$. However, for two events, E_1, E_2 it is not necessary that $\operatorname{Tr} \rho(E_1 \vee E_2) \leqslant \operatorname{Tr} \rho E_1 + \operatorname{Tr} \rho E_2$. In short, subadditivity property for probability need not hold good. But this property is retained whenever E_1 and E_2 commute with each other.

If ρ is a state and X is an element of $\mathcal{O}(\mathcal{H})$ with spectral resolution (2.1) then $\operatorname{Tr} \rho E_\lambda$ is the probability that X takes the value λ in the state ρ whenever $\lambda \in \sigma(X)$. Thus the *expectation* of X in the state ρ is given by

$$\sum_{\lambda \in \sigma(X)} \lambda \operatorname{Tr} \rho E_\lambda = \operatorname{Tr} \rho \sum_{\lambda \in \sigma(X)} \lambda E_\lambda = \operatorname{Tr} \rho X.$$

More generally, the expectation of $\varphi(X)$ defined by (2.4) is given by $\operatorname{Tr} \rho \varphi(X)$. In particular, the *variance* of X in the state ρ, denoted by $\operatorname{Var}(X|\rho)$ is given by

$$\begin{aligned} \operatorname{Var}(X|\rho) &= \operatorname{Tr} \rho X^2 - (\operatorname{Tr} \rho X)^2 \\ &= \operatorname{Tr} \rho(X - m)^2 \end{aligned}$$

where $m = \operatorname{Tr} \rho X$ is the expectation or *mean* of X in the state ρ. This shows, in particular, that $\operatorname{Var}(X|\rho)$ vanishes if and only if the restriction of X to the range of ρ is a scalar multiple of the identity.

The spectral theorem implies that the extreme points of the convex set $\mathcal{S}(\mathcal{H})$ are one dimensional projections of the form $|\psi\rangle\langle\psi|$ where $|\psi\rangle$ is a unit vector in $\mathcal{H}$. Here, the projection remains unaltered if $|\psi\rangle$ is replaced by $c|\psi\rangle$ where c is a scalar of modulus unity. Extreme points of $\mathcal{S}(\mathcal{H})$ are called *pure states* and a pure state is a one dimensional projection which, in turn, is determined by a unit vector in $\mathcal{H}$ modulo a scalar of modulus unity. By abuse of language any determining unit vector itself is called a pure state. Thus whenever we say that a unit vector $|\psi\rangle$ is a pure state we mean the density operator $|\psi\rangle\langle\psi|$. By spectral theorem any state ρ can be expressed as $\sum_j p_j |\psi_j\rangle\langle\psi_j|$ where $p_1, p_2, \ldots$ is a finite probability distribution and $\{|\psi_j\rangle, j = 1, 2, \ldots\}$ is an orthonormal set of vectors in $\mathcal{H}$. If $\{|\psi_j\rangle\}$ is any set of unit vectors and $p_j, j = 1, 2, \ldots$ is a probability distribution then $\sum_j p_j |\psi_j\rangle\langle\psi_j|$ is a state. If $|\psi\rangle$ is a pure state and X is a real-valued observable then its variance $\operatorname{Var}(X\||\psi\rangle)$ in the pure state $|\psi\rangle$ is zero if and only if $|\psi\rangle$ is an eigenvector for X. Thus, even in a pure state $|\psi\rangle$, an observable need not have a degenerate distribution. This is a significant departure from classical probability.

Hereafter, unless otherwise explicitly mentioned, we shall mean by an observable a real-valued observable. Let X, Y be two observables, ρ a state and let $m = \operatorname{Tr} \rho X$, $m' = \operatorname{Tr} \rho Y$ their respective means. Put $\widetilde{X} = X - m$, $\widetilde{Y} = Y - m'$ and consider the nonnegative function

$$f(z) = \operatorname{Tr} \rho(\widetilde{X} + z\widetilde{Y})^\dagger(\widetilde{X} + z\widetilde{Y}), \quad z \in \mathbb{C}.$$

Then the inequality

$$\inf_{z \in \mathbb{C}} f(z) \geqslant 0$$

implies (see [6], [16])

$$\operatorname{Var}(X|\rho)\operatorname{Var}(Y|\rho) \geqslant \left\{ \operatorname{Tr} \rho \frac{1}{2i}(\widetilde{X}\widetilde{Y} - \widetilde{Y}\widetilde{X}) \right\}^2 + \left\{ \operatorname{Tr} \rho \frac{1}{2}(\widetilde{X}\widetilde{Y} - \widetilde{Y}\widetilde{X}) \right\}^2 \quad (2.5)$$

and thus puts a lower bound on the product of the variances of X and Y in a state ρ. The quantum probability of finite level systems we have described here has a natural generalization when $\mathcal{H}$ is an infinite dimensional Hilbert space. When $\mathcal{H} = L^2(\mathbb{R})$, $X = q$, $Y = p$ are the well-known position and momentum operators satisfying the Heisenberg commutation relations $qp - pq = i\hbar$ the inequality (2.5) yields the special form

$$\operatorname{Var}(q||\psi\rangle)\operatorname{Var}(p||\psi\rangle) \geqslant \frac{\hbar^2}{4} \quad \forall \quad |\psi\rangle \in \mathcal{D}$$

where $\mathcal{D}$ is a dense domain in $\mathcal{H}$ where unbounded operators like qp, pq etc. are well-defined. Thus (2.5) is at the heart of the Heisenberg's principle of uncertainty.

Now we introduce a notion which is more general than that of an observable. Indeed, it plays an important role in the quantum version of Shannon's coding theorems of classical information theory.

Definition 2.1. *A generalized measurement $\mathcal{L}$ of a finite level quantum system with Hilbert space $\mathcal{H}$ is a pair (S, L) where S is a finite set and $L : S \to \mathcal{B}(\mathcal{H})$ is a map satisfying the condition:*

$$\sum_{s \in S} L(s)^\dagger L(s) = I. \tag{2.6}$$

Such a generalized measurement $\mathcal{L} = (S, L)$ has the following interpretation. If the system is in the state ρ and the measurement $\mathcal{L}$ is performed then the 'value' $s \in S$ is obtained with probability $\operatorname{Tr} L(s)\rho L(s)^\dagger$ and the system 'collapses' to a new state

$$\frac{L(s)\rho L(s)^\dagger}{\operatorname{Tr} L(s)\rho L(s)^\dagger}. \tag{2.7}$$

If, for example, the system is initially in the state ρ, a generalized measurement $\mathcal{L}_1 = (S_1, L_1)$ is performed and followed by another generalized measurement $\mathcal{L}_2 = (S_2, L_2)$ then the probability of obtaining the value $s_1 \in S_1$ is $\operatorname{Tr} L_1(s_1)\rho L_1(s_1)^\dagger$ and the conditional probability of getting the value $s_2 \in S_2$ from $\mathcal{L}_2$ given the value s_1 from $\mathcal{L}_1$ is

$$\operatorname{Tr} L_2(s_2) \left\{ \frac{L_1(s_1)\rho L_1(s_1)^\dagger}{\operatorname{Tr} L_1(s_1)\rho L_1(s_1)^\dagger} \right\} L_2(s_2)^\dagger.$$

Thus the probability of obtaining the value (s_1, s_2) from $\mathcal{L}_1$ followed by $\mathcal{L}_2$ is equal to

$$p(s_1, s_2) = \operatorname{Tr} L_2(s_2)L_1(s_1)\rho \, L_1(s_1)^\dagger L_2(s_2)^\dagger$$

and the final collapsed state is

$$\frac{L_2(s_2)L_1(s_1)\rho L_1(s_1)^{\dagger}L_2(s_2)^{\dagger}}{p(s_1, s_2)}.$$

More generally, if the measurements $\mathcal{L}_i = (S_i, L_i)$, $i = 1, 2, \ldots, m$ are performed in succession on a quantum system with initial state ρ then the probability $p(s_1, s_2, \ldots, s_m)$ of getting the sequence $s_1, s_2, \ldots, s_m$ of values $s_j \in S_j \forall j$ is given by

$$p(s_1, s_2, \ldots, s_m) = \text{Tr}\, L_m(s_m)L_{m-1}(s_{m-1}) \cdots L_1(s_1)\rho L_1(s_1)^{\dagger}L_2(s_2)^{\dagger} \cdots L_m(s_m)^{\dagger}$$

and the final collapsed state is

$$\frac{1}{p(s_1, s_2, \ldots, s_m)} L_m(s_m)L_{m-1}(s_{m-1}) \ldots L_1(s_1)\rho L_1(s_1)^{\dagger}L_2(s_2)^{\dagger} \ldots L_m(s_m)^{\dagger}.$$

This at once suggests the product rule for measurements $\mathcal{L}_i = (S_i, L_i)$ $i = 1, 2$ as $\mathcal{L} = (S_1 \times S_2, \widetilde{L})$ where

$$\widetilde{L}(s_1, s_2) = L_2(s_2)L_1(s_1), \quad s_1 \in S_1, s_2 \in S_2.$$

The measurement $\mathcal{L}$ stands for the measurement $\mathcal{L}_1$ followed by the measurement $\mathcal{L}_2$.

If $\mathcal{L} = (S, L)$ is a measurement with $S \subset \mathbb{R}$ or $\mathbb{C}$ then its *expectation* in the state ρ is given by

$$\sum_{s \in S} s\, \text{Tr}\, L(s)\rho L(s)^{\dagger} = \sum_{s \in S} s\, \text{Tr}\, \rho L(s)^{\dagger}L(s).$$

If $S \subset \mathbb{R}$ its variance $\text{Var}(\mathcal{L}|\rho)$ in the state ρ is given by

$$\sum_{s \in S} s^2\, \text{Tr}\, \rho\, L(s)^{\dagger}L(s) - \left(\sum_{s \in s} s\, \text{Tr}\, \rho L(s)^{\dagger}L(s)\right)^2.$$

When $L(s)$ is a projection for every $s \in S$ then (S, L) is called a *projective* or *von Neumann measurement*. If, in addition, $S \subset \mathbb{R}$ then the hermitian operator $\sum_{s \in S} s\, L(s)$ is an observable and our notion of generalized measurement reduces to measuring an observable. It may be of some interest to formulate and obtain an uncertainty principle for a pair of two real-valued measurements.

For a measurement with values in an abstract set S it is natural to replace the notion of variance by its entropy in a state ρ. Thus we consider the quantity

$$H(\mathcal{L}|\rho) = -\sum_{s \in S} p(s)\, \log_2 p(s) \tag{2.8}$$

where

$$p(s) = \text{Tr}\, \rho\, L(s)^{\dagger}\, L(s)$$

and call it the *entropy* of the measurement $\mathcal{L} = (S, L)$ in the state ρ. With this definition one has the following entropic uncertainty principle.

Theorem 2.2 ([11], [12]). *Let $\mathcal{L} = (S, L)$, $\mathcal{M} = (T, M)$ be two generalized measurements of a finite level quantum system in a Hilbert space $\mathcal{H}$. Let $L(s)^\dagger L(s) = X(s)$, $M(t)^\dagger M(t) = Y(t)$, $s \in S$, $t \in T$. Then for any state ρ the following holds:*

$$H(\mathcal{L}|\rho) + H(\mathcal{M}|\rho) \geqslant -2 \log_2 \max_{s,t} \left\| X(s)^{1/2} Y(t)^{1/2} \right\|. \tag{2.9}$$

Remark   It is important to note that the right hand side in the inequality (2.9) is independent of ρ.

If X_i, $1 \leqslant i \leqslant k$ are k observables, ρ is a state in $\mathcal{H}$ and $\operatorname{Tr} \rho X_i = m_i$ define the scalar

$$\nu_{ij} = \frac{1}{2} \operatorname{Tr} \rho \left\{ (X_i - m_i)(X_j - m_j) + (X_j - m_j)(X_i - m_i) \right\}. \tag{2.10}$$

Then the real symmetric matrix $((\nu_{ij}))$ of order k is called the *covariance matrix* of the observables $X_1, X_2, \ldots, X_k$ in the state ρ and denoted by $\operatorname{Cov}\left(X_1, X_2, \ldots, X_k \middle| \rho\right)$. It is a positive semidefinite matrix and it is important to note the symmetrization in i, j in the right hand side of (2.10). Without such a symmetrization ν_{ij} could be a complex scalar.

Till now we talked about a single quantum system. Suppose we have a composite quantum system made out of several simple systems $A_1, A_2, \ldots, A_k$ with respective Hilbert spaces $\mathcal{H}_{A_1}, \mathcal{H}_{A_2}, \ldots, \mathcal{H}_{A_k}$. Then the Hilbert space of the joint system $A_1 A_2 \ldots A_k$ is the tensor product

$$\mathcal{H}_{A_1 \ldots A_k} = \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2} \otimes \cdots \otimes \mathcal{H}_{A_k}.$$

This is the quantum probabilistic analogue of cartesian product of sample spaces in classical probability. It is clear that

$$\dim \mathcal{H}_{A_1 \ldots A_k} = \prod_{i=1}^{k} \dim \mathcal{H}_{A_i},$$

dim indicating dimension. If ρ_i is a state in $\mathcal{H}_{A_i} \, \forall \, i$ then $\rho_1 \otimes \cdots \otimes \rho_k$ is a state of the composite system $A_1 A_2 \ldots A_k$ called the *product* state. If ρ is a state in $\mathcal{H}_{A_1 \ldots A_k}$ and we take its relative trace over $\mathcal{H}_{A_{i_1}}, \mathcal{H}_{A_{i_2}}, \ldots, \mathcal{H}_{A_{i_\ell}}$ then we get the marginal state of the system $A_{r_1}, A_{r_2} \ldots, A_{r_m}$ where $\{1, 2, \ldots, k\}$ is the disjoint union $\{i_1, i_2, \ldots, i_\ell\} \cup \{r_1, r_2, \ldots, r_m\}$ with $\ell + m = k$. In this context of composite quantum systems there arises a new distinguishing feature of the subject with a remarkable role in physics as well as information theory. It is the existence of a very rich class of states in $\mathcal{H}_{A_1 A_2 \ldots A_k}$ which do not belong to the convex hull of all product states. Such states are called *entangled states* and they constitute a rich resource in quantum communication [14].

Till now we restricted ourselves to quantum probability. Now we describe a few basic concepts in quantum statistics dealing with a parametric family of quantum states of a finite level system. Let Γ be a *parameter space* and let $\{\rho(\theta), \theta \in \Gamma\}$ be a *parametric family* of states in a Hilbert space $\mathcal{H}$. Suppose X is an observable, i.e., an element of $\mathcal{O}(\mathcal{H})$ and

$$\operatorname{Tr} \rho(\theta) X = f(\theta), \ \theta \in \Gamma, \tag{2.11}$$

where f is a real-valued function on Γ. then we say that the observable X is an *unbiased estimator* of the *parametric function f* on Γ.

When the parametric family $\{\rho(\theta), \theta \in \Gamma\}$ is fixed we write

$$\mathrm{Var}(X|\theta) = \mathrm{Var}(X|\rho(\theta)) \tag{2.12}$$

If $X_1, X_2, \ldots, X_m$ are m observables we write

$$\mathrm{Cov}(X_1, \ldots, X_m|\theta) = \mathrm{Cov}(X_1, \ldots, X_m|\rho(\theta)). \tag{2.13}$$

A real-valued function f on Γ is said to be *estimable* with respect to $\{\rho(\theta), \theta \in \Gamma\}$ if there exists an observable $X \in \mathcal{O}(\mathcal{H})$ such that

$$\mathrm{Tr}\,\rho(\theta)X = f(\theta) \;\forall\; \theta \in \Gamma.$$

Denote by $\mathcal{E}(\Gamma)$ the real linear space of all such estimable functions. An observable X is said to be *balanced* with respect to the family $\{\rho(\theta), \theta \in \Gamma\}$ if $\mathrm{Tr}\,\rho(\theta)X = 0 \;\forall\; \theta \in \Gamma$. Denote by $\mathcal{N}$ the real linear space of all such balanced observables. For any $f \in \mathcal{E}(\Gamma)$, an unbiased estimator X of f write

$$\nu_f(\theta) = \inf\left\{\mathrm{Var}(X + Z|\theta),\; Z \in \mathcal{N}\right\}.$$

It is natural to look for good lower bounds for the function $\nu_f(\theta)$. We shall examine this problem in the next section and study some examples. If $f_j, 1 \leqslant j \leqslant m$ are estimable parametric functions we shall also look for matrix lower bounds for the positive semidefinite matrices $\mathrm{Cov}(X_1, \ldots, X_m|\theta)$ as each X_i varies over all unbiased estimators of f_i for each $i = 1, 2, \ldots, m$.

For a more detailed introduction to quantum probability theory we refer to [15], [16]. For an initiation to estimation theory and testing hypotheses in quantum statistics we refer to [7], [9], [10], References [7], [10], [14], [17] contain applications of the theory of generalized measurements.

3. The Fisher information form and the Cramér-Rao-Bhattacharya tensor

We consider a fixed parametric family $\{\rho(\theta), \theta \in \Gamma\}$ of states of a finite level quantum system in a Hilbert space $\mathcal{H}$ with parameter space Γ. As mentioned in the preceding section denote by $\mathcal{E}(\Gamma)$ and $\mathcal{N}$ respectively the real linear spaces of estimable functions and balanced observables. Recall that for any two unbiased estimators X and Y of an element $f \in \mathcal{E}(\Gamma)$, the observable $X - Y$ is an element of $\mathcal{N}$.

Definition 3.1. *A map $F : \Gamma \to \mathcal{B}(\mathcal{H})$ is called a Fisher map for the family $\{\rho(\theta), \theta \in \Gamma\}$ of states in $\mathcal{H}$ if the following two conditions hold:*

(i) $\mathrm{Tr}\,\rho(\theta)F(\theta) = 0 \quad \forall\; \theta \in \Gamma$,
(ii) $\mathrm{Tr}\,\rho(\theta)\left\{F(\theta)^\dagger X + XF(\theta)\right\} = 0 \quad \forall\; \theta \in \Gamma, X \in \mathcal{N}$.

Denote by $\mathcal{F}$ the real linear space of all Fisher maps with respect to $\{\rho(\theta), \theta \in \Gamma\}$ and by $\mathcal{A}(\Gamma)$ the algebra of all real-valued functions on Γ. If $a \in \mathcal{A}(\Gamma)$ and $F \in \mathcal{F}$

then aF defined by $(aF)(\theta) = a(\theta)F(\theta)$ is also in $\mathcal{F}$. In other words $\mathcal{F}$ is an $\mathcal{A}(\Gamma)$-module. For any two Fisher maps F, G in $\mathcal{F}$ define

$$
\begin{aligned}
\mathcal{I}(F, G)(\theta) &= \operatorname{Tr} \rho(\theta) \frac{1}{2} \left(F(\theta)^\dagger G(\theta) + G(\theta)^\dagger F(\theta) \right) \\
&= \operatorname{Re} \operatorname{Tr} \rho(\theta) F(\theta)^\dagger G(\theta).
\end{aligned}
\tag{3.1}
$$

Then $\mathcal{I}$ is called the *Fisher information form* associated with $\{\rho(\theta), \theta \in \Gamma\}$. It may be noted that, for all $F, F_1, F_2, G \in \mathcal{F}$ and $a \in \mathcal{A}(\Gamma)$, the following hold:

$$
\begin{aligned}
\mathcal{I}(F, G) &= \mathcal{I}(G, F), \\
\mathcal{I}(aF, G) &= a\,\mathcal{I}(F, G), \\
\mathcal{I}(F_1 + F_2, G) &= \mathcal{I}(F_1, G) + \mathcal{I}(F_2, G), \\
\mathcal{I}(F, F) &\geqslant 0.
\end{aligned}
$$

In particular, for any F_i, $1 \leqslant i \leqslant n$ in $\mathcal{F}$ the matrix

$$
\mathcal{I}_n(F_1, F_2, \ldots, F_n | \theta) = \left((\mathcal{I}(F_i, F_j)(\theta)) \right), \theta \in \Gamma, i, j \in \{1, 2, \ldots, n\}
\tag{3.2}
$$

is positive semidefinite. It is called the *information matrix* at θ corresponding to the elements F_i, $1 \leqslant i \leqslant n$ in $\mathcal{F}$.

If $f \in \mathcal{E}(\Gamma)$, $F \in \mathcal{F}$ define

$$
\lambda(f, F)(\theta) = \operatorname{Tr} \rho(\theta) \frac{1}{2} \left(F(\theta)^\dagger X + X F(\theta) \right), \quad \theta \in \Gamma
\tag{3.3}
$$

where X is any unbiased estimator of f. Note that, in view of property (ii) in Definition 3.1 the right hand side of (3.3) is independent of the choice of the unbiased estimator of f. Clearly, $\lambda(f, F)$ is real linear in the variable f when F is fixed and $\mathcal{A}(\Gamma)$-linear in the variable F when f is fixed. Thus $\lambda(\cdot, \cdot)$ can be viewed as an element of $\mathcal{E}(\Gamma) \otimes \mathcal{F}$. We call $\lambda(\cdot, \cdot)$ the *Cramér-Rao-Bhattacharya tensor* or simply the *CRB-tensor* associated with $\{\rho(\theta), \theta \in \Gamma\}$.

Let $f_i \in \mathcal{E}(\Gamma)$, X_i an unbiased estimator of f_i for each $1 \leqslant i \leqslant m$ and let F_j, $1 \leqslant j \leqslant n$ be Fisher maps with respect to $\{\rho(\theta), \theta \in \Gamma\}$. Define the $m \times m$ matrix

$$
\begin{aligned}
\Lambda_{mn}(\theta) &= ((\lambda_{ij}(\theta))), \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n, \quad \theta \in \Gamma \tag{3.4} \\
\lambda_{ij}(\theta) &= \lambda(f_i, F_j)(\theta) \quad \theta \in \Gamma, \tag{3.5}
\end{aligned}
$$

λ being the CRB-tensor. We now introduce the family of positive semidefinite sesquilinear forms indexed by $\theta \in \Gamma$ in the vector space $\mathcal{B}(\mathcal{H})$ by

$$
B_\theta(X, Y) = \operatorname{Tr} X^\dagger \rho(\theta) Y, \quad X, Y \in \mathcal{B}(\mathcal{H}).
\tag{3.6}
$$

By property (i) in Definition 3.1, equations (3.3) and (3.5) we have

$$
\begin{aligned}
\lambda_{ij}(\theta) &= \operatorname{Tr} \rho(\theta) \frac{1}{2} \left\{ F_j(\theta)^\dagger (X_i - f_i(\theta)) + (X_i - f_i(\theta)) F_j(\theta) \right\} \\
&= \operatorname{Re} B_\theta \left(X_i - f_i(\theta), F_j(\theta)^\dagger \right).
\end{aligned}
$$

Multiplying both sides by the real scalars $a_i b_j$ and adding over $1 \leqslant i \leqslant m$, $1 \leqslant i \leqslant n$, we obtain

$$\mathbf{a}'\Lambda_{mn}(\theta)\mathbf{b} = \mathrm{Re}\, B_\theta \left(\sum_{i=1}^{m} a_i(X_i - f_i(\theta)), \sum_{j=1}^{n} b_j F_j(\theta)^\dagger \right) \qquad (3.7)$$

where Λ_{mn} and B_θ are given by (3.4), (3.5) and (3.6) and $\mathbf{a}, \mathbf{b}$ are respectively column vectors of length m, n with prime $'$ indicating transpose. Now an application of the Cauchy-Schwarz inequality to the right hand side of (3.7) implies

$$
\begin{aligned}
\left(\mathbf{a}'\Lambda_{mn}(\boldsymbol{\theta})\,\mathbf{b}\right)^2 \;\leqslant\; & B_\theta \left(\sum_{i=1}^{n} a_i(X_i - f_i(\theta)), \sum_{i=1}^{m} a_i(X_i - f_i(\theta)) \right) \\
& \times B_\theta \left(\sum_{j=1}^{n} b_j F_j(\theta)^\dagger, \sum_{j=1}^{n} b_j F_j(\theta)^\dagger \right) \\
= \; & \left\{ \mathbf{a}'\mathrm{Cov}\left(X_1, X_2, \ldots, X_m | \theta\right)\mathbf{a} \right\} \left\{ \mathbf{b}'\mathcal{I}_n\left(F_1, F_2, \ldots, F_n | \theta\right)\mathbf{b} \right\}.
\end{aligned}
$$

Dividing both sides of this inequality by $\mathbf{b}'\mathcal{I}_n\left(F_1, F_2, \ldots, F_n | \theta\right)\mathbf{b}$, fixing $\mathbf{a}$ and maximizing the left hand side over all $\mathbf{b}$ satisfying $\mathcal{I}_n\left(F_1, F_2, \ldots, F_n | \theta\right)\mathbf{b} \neq \mathbf{0}$ we obtain the matrix inequality:

$$\Lambda_{mn}(\theta)\mathcal{I}_n^-(F_1, F_2, \ldots, F_n | \theta)\Lambda_{mn}(\theta)' \leqslant \mathrm{Cov}(X_1, X_2, \ldots X_m | \theta),$$

$\mathcal{I}_n^-$ denoting the generalized inverse of $\mathcal{I}_n(F_1, F_2, \ldots, F_n | \theta)$. In other words we have proved the following theorem

Theorem 3.1 (Quantum Cramér-Rao-Bhattacharya (CRB) inequality). *Let* $\{\rho(\theta), \theta \in \Gamma\}$ *be a parametric family of states of a finite level quantum system in a Hilbert space* $\mathcal{H}$, f_i, $1 \leqslant i \leqslant m$ *estimable functions on* Γ, X_i *an unbiased estimator of* f_i *for each* i *and let* F_j, $1 \leqslant j \leqslant n$ *be Fisher maps with respect to* $\{\rho(\theta), \theta \in \Gamma\}$. *Then the following matrix inequality holds:*

$$\mathrm{Cov}\left(X_1, X_2, \ldots, X_m | \theta\right) \geqslant \Lambda_{mn}(\theta)\mathcal{I}_n^-\left(F_1, F_2, \ldots, F_n | \theta\right)\Lambda_{mn}(\theta)' \; \forall\, \theta \in \Gamma$$

where $\Lambda_{mn}(\theta)$ *is the* $m \times n$ *matrix defined by* (3.3)-(3.5) *and* $\mathcal{I}_n^-(F_1, F_2, \ldots, F_n | \theta)$ *is the generalized inverse of the Fisher information matrix* $\mathcal{I}_n(F_1, F_2, \ldots, F_n | \theta)$ *associated with* $F_1, F_2, \ldots, F_n$.

Proof. Immediate.   ∎

Corollary 3.1. *Let* X_i, $1 \leqslant i \leqslant m$, F_j, $1 \leqslant j \leqslant n$ *be as in Theorem 3.1. Then*

$$
\begin{aligned}
& \Lambda_{mn}(\theta)\mathcal{I}_n^-\left(F_1, F_2, \ldots, F_n | \theta\right)\Lambda_{mn}(\theta)' \\
& \quad \geqslant\; \Lambda_{mn-1}(\theta)\mathcal{I}_{n-1}^-(F_1, F_2, \ldots, F_{n-1} | \theta)\Lambda_{mn-1}(\theta)', \theta \in \Gamma
\end{aligned}
$$

for $n \geqslant 2$.

Proof. This is immediate from the fact that both the sides of the inequality above are arrived at by taking supremum over certain sets in $\mathbb{R}^n$, the set for the left hand side being larger than the set for the right hand side. ■

We call the right hand side of the inequality in Theorem 3.1 the CRB lower bound.

Remark 1   Theorem 3.1 and Corollary 3.1 imply the possibility of improving the CRB lower bound by searching for a larger class of $\mathcal{A}(\Gamma)$-linearly independent Fisher maps for a parametric family of states.

Remark 2   The CRB lower bound in Theorem 3.1 has some natural invariance properties. If f_i, $1 \leqslant i \leqslant m$ are fixed and X_i, $F_i(\theta)$, $\rho(\theta)$ are changed respectively to $UX_iU^\dagger$, $UF_i(\theta)U^\dagger$, $U\rho(\theta)U^\dagger$ by a fixed unitary operator U in $\mathcal{H}$ then the CRB lower bound in Theorem 3.1 remains the same.

If the Fisher maps F_j are replaced by

$$G_j(\theta) = \sum_{r=1}^{n} \alpha_{jr}(\theta) F_r(\theta), \quad 1 \leqslant j \leqslant n \tag{3.8}$$

where the matrix $A(\theta) = ((\alpha_{rs}(\theta)))$ is invertible for all θ then

$$\Lambda_{mn}(\theta)\mathcal{I}_n^-(F_1,\ldots,F_n|\theta)\Lambda_{mn}(\theta)' = \tilde{\Lambda}_{mn}(\theta)\mathcal{I}_n^-(G_1,\ldots,G_n|\theta)\tilde{\Lambda}_{mn}(\theta)',$$

the tilde over Λ_{mn} on the right hand side indicating that G_i's are used in place of F_i's. In other words the CRB bound is invariant under $\mathcal{A}(\Gamma)$-linear invertible transformations of the form (3.8).

Example 3.1 Let $\Gamma = (a,b)$, $\mathcal{H} = \mathbb{C}^n$ and let

$$\rho(\theta) = \operatorname{diag}\left(p_1(\theta), p_2(\theta), \ldots, p_n(\theta)\right), \quad \theta \in \Gamma$$

be states in $\mathbb{C}_n$ with respect to the standard orthonormal basis, diag denoting diagonal matrix. An estimable function f on Γ has the form

$$f(\theta) = \sum_{i=1}^{n} a_i \, p_i(\theta)$$

where a_i are real scalars. An unbiased estimator X for f is

$$X = \operatorname{diag}(a_1, a_2, \ldots, a_n).$$

Note that $p_i(\theta) \geqslant 0$ and $\sum_i p_i(\theta) = 1 \; \forall \; \theta \in \Gamma$. Assume that $p_i(\theta)$ are differentiable in θ and $p_i(\theta) > 0 \; \forall \; i, \theta$. Then

$$F(\theta) = \operatorname{diag}\left(\frac{p_1'(\theta)}{p_1(\theta)}, \frac{p_2'(\theta)}{p_2(\theta)}, \ldots, \frac{p_n'(\theta)}{p_n(\theta)}\right)$$

yields a Fisher map with

$$\mathcal{I}(F,F)(\theta) = \sum_{i=1}^{n} \frac{p_i'(\theta)^2}{p_i(\theta)}$$

and

$$\lambda(f, F) = \sum_{i=1}^{n} a_i \, p_i'(\theta) = f'(\theta).$$

Theorem 3.1 for the single observable X and single Fisher map yields

$$\text{Var}\,(Y|\theta) \geqslant \frac{\left(\sum_{i=1}^{n} a_i \, p_i'(\theta)\right)^2}{\sum_{i=1}^{n} \frac{p_i'(\theta)^2}{p_i(\theta)}} \quad \forall\, \theta \in (a, b)$$

and any unbiased estimator Y of f. This is the Cramér-Rao inequality for finite sample spaces in classical mathematical statistics.

Example 3.2(Quantum version of Barankin's example [1], [21]). Let $\rho(\theta)$ be an invertible density operator for every θ in Γ. For any $\gamma \in \Gamma$ define

$$F_\gamma(\theta) = \rho(\gamma)\rho(\theta)^{-1} - 1.$$

Then F_γ is a Fisher map and for any estimable function $f \in \mathcal{E}(\Gamma)$ we have

$$\lambda(f, F_\gamma)(\theta) = f(\gamma) - f(\theta).$$

The Fisher information form $\mathcal{I}$ satisfies

$$\mathcal{I}(F_{\gamma_1}, F_{\gamma_2})(\theta) = \text{Re Tr}\, \rho(\gamma_1)\rho(\theta)^{-1}\rho(\gamma_2) - 1$$

If X is an unbiased estimate of $f \in \mathcal{E}(\Gamma)$ one obtains as a special case the CRB bound

$$\text{Var}(X|\theta) \geqslant (f(\gamma_1) - f(\theta), f(\gamma_2) - f(\theta), \ldots, f(\gamma_n) - f(\theta))$$
$$\mathcal{I}_n^-(\gamma_1, \gamma_2, \ldots, \gamma_n, \theta)\,(f(\gamma_1) - f(\theta), \ldots, f(\gamma_n) - f(\theta))'$$

where $\mathcal{I}_n^-(\gamma_1, \gamma_2, \ldots, \gamma_n\theta)$ is the generalized inverse of the information matrix

$$\left(\left(\text{Re Tr}\, \rho(\gamma_i)\rho(\theta)^{-1}\rho(\gamma_j) - 1\right)\right)$$

for any $\gamma_1, \gamma_2, \ldots, \gamma_n \in \Gamma$.

Example 3.3(Quantum Bhattacharya bound [2]). Let $\Gamma \subseteq \mathbb{R}^d$ be a connected open set and let $\rho(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Gamma$ be a family of invertible states such that the correspondence $\boldsymbol{\theta} \to \rho(\boldsymbol{\theta})$ is C^m-smooth. then every estimable function f is also C^m-smooth. For any linear differential operator D on Γ with C^m-coefficients satisfying $D\,1 = 0$ define

$$F_D(\boldsymbol{\theta}) = (D\rho)(\boldsymbol{\theta})\rho(\boldsymbol{\theta})^{-1}$$

where D is applied to every matrix entry of $\rho(\cdot)$ on the right hand side in some fixed orthonormal basis. Then F_D is a Fisher map and the CRB tensor λ satisfies

$$\lambda(f, F_D)(\boldsymbol{\theta}) = (Df)(\boldsymbol{\theta}) \quad \forall\, f \in \mathcal{E}(\Gamma).$$

If D_1, D_2 are two linear differential operators in Γ with C^m-coefficients annihilating the constant function 1 the Fisher information satisfies

$$\mathcal{I}(F_{D_1}, F_{D_2})(\boldsymbol{\theta}) = \text{Re Tr}\,(D_1\rho)(\boldsymbol{\theta})\,\rho(\boldsymbol{\theta})^{-1}(D_2\rho)(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Gamma.$$

If X is an unbiased estimate of f and D_i, $1 \leqslant i \leqslant n$ are C^m-differential operators on Γ then the CRB inequality has the form

$$\text{Var}(X|\theta) \geqslant (D_1 f, \ldots, D_n f)(\boldsymbol{\theta}) \mathcal{I}_n^-(D_1, D_2, \ldots, D_n|\boldsymbol{\theta})(D_1 f, \ldots, D_n f)(\boldsymbol{\theta})'$$

where $\mathcal{I}_n^-(D_1, D_2, \ldots, D_n|\boldsymbol{\theta})$ is the generalized inverse of the positive semidefinite matrix

$$\left(\left(\text{Re Tr } (D_i \rho)(\boldsymbol{\theta}) \rho(\boldsymbol{\theta})^{-1}(D_j \rho)(\boldsymbol{\theta})\right)\right), \quad i, j \in \{1, 2, \ldots, n\}.$$

Example 3.4 Example 3.2 leads us to the following natural abstraction. Suppose Γ is a d-dimensional C^m-manifold and $\boldsymbol{\theta} \to \rho(\boldsymbol{\theta})$ is a C^m-smooth parametrization of states in $\mathcal{H}$ as $\boldsymbol{\theta}$ varies in Γ. If L is a smooth vector field on Γ then

$$F_L(\theta) = (L\rho)(\boldsymbol{\theta})\rho(\boldsymbol{\theta})^{-1}, \quad \theta \in \Gamma$$

is a C^m-smooth Fisher map with respect to $\{\rho(\theta), \theta \in \Gamma\}$ under the assumption that $\rho(\theta)^{-1}$ exists for every θ. C^m-smooth Fisher maps constitute a $C^m(\Gamma)$-module and $\mathcal{E}(\Gamma) \subset C^m(\Gamma)$. The CRB tensor λ and the Fisher information form $\mathcal{I}$ satisfy the relations

$$\begin{aligned} \lambda(f, F_L)(\theta) &= (Lf)(\theta) \\ \mathcal{I}(F_L, F_M)(\theta) &= \text{Re Tr } (L\rho)(\theta)\rho(\theta)^{-1}(M\rho)(\theta) \end{aligned}$$

for any two vector fields L, M. As a special case of the CRB inequality we have for any unbiased estimator X of $f \in \mathcal{E}(\Gamma)$,

$$\text{Var}(X|\theta) \geqslant \frac{(Lf)(\theta)^2}{\text{Tr } \rho(\theta)^{-1}(L\rho)(\theta)^2}, \quad \theta \in \Gamma$$

for any C^m-smooth vector field L.

As a special case of the example above, consider a connected Lie group Γ with Lie algebra $\mathcal{G}$. Let

$$\rho(g) = U_g \, \rho_0 \, U_g^\dagger, \quad g \in \Gamma$$

where ρ_0 is a fixed invertible state. Any element L of $\mathcal{G}$ is looked upon as a left invariant vector field on Γ. let $U_{\exp t L} = \exp t \, \pi(L)$, $t \in \mathbb{R}$, $L \in \mathcal{G}$ where $L \to \pi(L)$ is a representation of $\mathcal{G}$ in $\mathcal{H}$. Then the CRB inequality yields

$$\text{Var}(X|g) \geqslant \frac{((Lf)(g))^2}{\text{Tr } \rho_0^{-1} [\pi(L), \rho_0]^2} \quad \forall \quad L \in \mathcal{G} \tag{3.9}$$

where X is an unbiased estimator of f. If L_i, $1 \leqslant i \leqslant d$ is a basis for $\mathcal{G}$ and the nonnegative definite matrix $\mathcal{I}_d$ is defined by

$$\mathcal{I}_d = \left(\left(\text{Re Tr } \rho_0^{-1} [\pi(L_i), \rho_0] [\pi(L_j), \rho_0]\right)\right), \quad i, j \in \{1, 2, \ldots, d\}$$

then a maximization over all elements L in $\mathcal{G}$ on the right hand side of (3.9) yields

$$\text{Var}(X|g) \geqslant (L_1 f, L_2 f, \ldots, L_d f)(g) \mathcal{I}_d^-(L_1 f, L_2 f, \ldots, L_d f)(g)',$$

$\mathcal{I}_d^-$ being the generalized inverse of $\mathcal{I}_d$.

Example 3.5(adapted from [3]). We now consider an example in which the different states $\rho(\theta)$ may fail to have an inverse, indeed, their ranges need not be the

same. Let $\Gamma \subseteq \mathbb{R}^d$ be an open domain and let $\rho(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Gamma$ obey the set of linear partial differential equations of the form

$$\frac{\partial \rho}{\partial \theta_j} = \frac{1}{2}\left(L_j(\boldsymbol{\theta})\,\rho(\boldsymbol{\theta}) + \rho(\boldsymbol{\theta})L_j(\boldsymbol{\theta})^\dagger\right), \quad 1 \leqslant j \leqslant d \qquad (3.10)$$

where the operators $L_j(\boldsymbol{\theta}) \in \mathcal{B}(\mathcal{H})$. Taking trace on both sides we see that

$$\mathrm{Re}\ \mathrm{Tr}\ \rho(\boldsymbol{\theta})L_j(\boldsymbol{\theta}) = 0, \quad 1 \leqslant j \leqslant d, \quad \boldsymbol{\theta} \in \Gamma.$$

If $\mathcal{I}m\ \mathrm{Tr}\ \rho(\boldsymbol{\theta})L_j(\boldsymbol{\theta}) = m_j(\boldsymbol{\theta})$ we can replace in (3.10) $L_j(\boldsymbol{\theta})$ by $L_j(\boldsymbol{\theta}) - im_j(\boldsymbol{\theta})$ without altering the differential equations. Hence we may assume, without loss of generality, that in (3.10)

$$\mathrm{Tr}\ \rho(\boldsymbol{\theta})L_j(\boldsymbol{\theta}) = 0, \quad 1 \leqslant j \leqslant d, \quad \boldsymbol{\theta} \in \Gamma. \qquad (3.11)$$

We then say that the states $\rho(\boldsymbol{\theta})$ which obey (3.10) and (3.11) constitute a *Liapunov family*.

A special case of such a Liapunov family of states is obtained when $d = 1$ and

$$\rho(\theta) = p(\theta)e^{\frac{1}{2}\theta L}\,\rho_0\,e^{\frac{1}{2}\theta L^\dagger}, \quad \theta \in \mathbb{R}$$

where L is a fixed operator in $\mathcal{H}$, ρ_0 is a fixed state and

$$p(\theta) = \left\{\mathrm{Tr}\ \rho_0\ e^{\frac{1}{2}\theta L^\dagger}\ e^{\frac{1}{2}\theta L}\right\}^{-1}.$$

Then

$$\rho'(\theta) = \frac{1}{2}\left\{\left(\frac{p'(\theta)}{p(\theta)} + L\right)\rho(\theta) + \rho(\theta)\left(\frac{p'(\theta)}{p(\theta)} + L\right)^\dagger\right\}.$$

If $\rho_0 = |\psi_0\rangle\langle\psi_0|$ is a pure state then every $\rho(\theta)$ is a pure state. Thus rank $\rho(\theta) = 1$ $\forall\ \theta \in \mathbb{R}$ and we have a situation where $\{\rho(\theta)\}$ admits a 'score operator function' with a classical part p'/p and a quantum part L.

Going back to the Liapunov family satisfying (3.10) and (3.11) we observe that each of the maps $\boldsymbol{\theta} \to L_j(\boldsymbol{\theta})$, $1 \leqslant j \leqslant d$ is a Fisher map. Indeed, if X is a balanced observable we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial \theta_j}(\mathrm{Tr}\,\rho(\boldsymbol{\theta})X) \\
&= \frac{1}{2}\,\mathrm{Tr}\,\left(L_j(\boldsymbol{\theta})\rho(\boldsymbol{\theta})X + \rho(\boldsymbol{\theta})L_j(\boldsymbol{\theta})^\dagger X\right) \\
&= \frac{1}{2}\,\mathrm{Tr}\,\rho(\theta)\left(L_j(\theta)^\dagger X + X L_j(\boldsymbol{\theta})\right).
\end{aligned}$$

For any estimable function f

$$\lambda(f, L_j)(\boldsymbol{\theta}) = \frac{\partial f}{\partial \theta_j}$$

and the Fisher information form $\mathcal{I}$ satisfies

$$\mathcal{I}(L_i, L_j)(\boldsymbol{\theta}) = \mathrm{Re}\ \mathrm{Tr}\ \rho(\boldsymbol{\theta})L_i(\boldsymbol{\theta})^\dagger L_j(\boldsymbol{\theta}).$$

If we write

$$\mathcal{I}_d(\boldsymbol{\theta}) = ((\mathcal{I}(L_i, L_j)(\boldsymbol{\theta}))), \quad i, j \in \{1, 2, \ldots, d\}$$

then the CRB inequality assumes the form

$$\text{Var}(X|\boldsymbol{\theta}) \geqslant (\nabla f)(\boldsymbol{\theta})\mathcal{I}_d^-(\boldsymbol{\theta})(\nabla f)(\boldsymbol{\theta})'$$

for any unbiased estimator X of f, ∇f being the gradient vector $\left(\frac{\partial f}{\partial \theta}, \frac{\partial f}{\partial \theta_2}, \ldots, \frac{\partial f}{\partial \theta_d}\right)$.

In the special case $d = 1$ introduced in the course of the discussion above the CRB bound assumes the form

$$\text{Var}(X|\theta) \geqslant \frac{(f'(\theta))^2}{\text{Tr } \rho(\theta)\left(\frac{p'(\theta)}{p(\theta)} + L\right)^\dagger \left(\frac{p'(\theta)}{p(\theta)} + L\right)}.$$

If $\rho(\boldsymbol{\theta})$, $\sigma(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Gamma$ are Liapunov families of states in Hilbert spaces $\mathcal{H}$, $\mathcal{K}$ respectively with coefficients $L_j(\boldsymbol{\theta})$, $M_j(\boldsymbol{\theta})$ in the respective differential equations corresponding to (3.10) then the tensor product states $\rho(\boldsymbol{\theta})\otimes\sigma(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Gamma$ constitute again a Liapunov family with the coefficients $L_j(\boldsymbol{\theta}) \otimes 1 + 1 \otimes M_j(\theta)$, $1 \leqslant j \leqslant d$ in the differential equations corresponding to (3.10) and its Fisher information form satisfies

$$\begin{aligned}\mathcal{I}\left(L_i \otimes 1 + 1 \otimes M_i, L_j \otimes 1 + 1 \otimes M_j\right)(\boldsymbol{\theta}) \\ = \quad \mathcal{I}(L_i, L_j)(\boldsymbol{\theta}) + \mathcal{I}(M_i, M_j)(\boldsymbol{\theta}).\end{aligned}$$

Eexample 3.6   Our last example is the case when $\rho(\boldsymbol{\theta})$ is a mixture of the form

$$\rho(\boldsymbol{\theta}) = \sum_{r=1}^{N} p_r(\boldsymbol{\theta})\, \rho_r(\boldsymbol{\theta})$$

where $\{p_r(\boldsymbol{\theta}), 1 \leqslant r \leqslant N\}$ is a family of probability distributions on the finite set $\{1, 2, \ldots, N\}$ indexed by $\boldsymbol{\theta} \in \Gamma$ and for each fixed r, $\{\rho_r(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Gamma\}$ is a Liapunov family of states obeying the differential equations

$$\frac{\partial \rho_r}{\partial \theta_j} = \frac{1}{2}\left\{L_{rj}(\boldsymbol{\theta})\, \rho_r(\boldsymbol{\theta}) + \rho_r(\boldsymbol{\theta})L_{rj}(\boldsymbol{\theta})^\dagger\right\}, \quad 1 \leqslant j \leqslant d, \ \boldsymbol{\theta} \in \Gamma$$

and the conditions

$$\text{Tr } \rho_r(\boldsymbol{\theta})\, L_{rj}(\boldsymbol{\theta}) = 0 \quad \forall \quad \boldsymbol{\theta} \in \Gamma.$$

Let now $f_i, 1 \leqslant i \leqslant m$ be estimable functions with respect to $\{\rho(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Gamma\}$ and let X_i be any unbiased estimator of f_i for each i. Differentiating with respect to θ_j the identity

$$\text{Tr } \rho(\boldsymbol{\theta})(X_i - f_i(\theta)) = 0$$

we get

$$\frac{\partial f_i}{\partial \theta_j} = \sum_{r=1}^{N} p_r(\boldsymbol{\theta})\, \text{Re Tr } M_{rj}(\boldsymbol{\theta})\rho_r(\boldsymbol{\theta})(X_i - f_i(\boldsymbol{\theta})) \tag{3.12}$$

where

$$M_{rj}(\boldsymbol{\theta}) = p_r(\boldsymbol{\theta})^{-1}\frac{\partial p_r}{\partial \theta_j} + L_{rj}(\boldsymbol{\theta}). \tag{3.13}$$

Multiplying both sides of (3.12) by real scalars $a_i b_j$ and adding over i and j we get

$$\mathbf{a}'\left(\left(\frac{\partial f_i}{\partial \theta_j}\right)\right)\mathbf{b} = \sum_{r=1}^{N} p_r(\boldsymbol{\theta})\, \mathrm{Tr}\,\left(\sum_{j=1}^{d} b_j M_{rj}(\boldsymbol{\theta})\right)\rho_r(\boldsymbol{\theta})\left(\sum_{i=1}^{m} a_i(X_i - f_i(\boldsymbol{\theta}))\right).$$

Applying Cauchy-Schwarz inequality to each trace scalar product on the right hand side followed by the same inequality to the scalar product with respect to the probability distribution $p_1(\boldsymbol{\theta}), p_2(\boldsymbol{\theta}), \ldots, p_N(\boldsymbol{\theta})$ we obtain

$$\left(\mathbf{a}'\left(\left(\frac{\partial f_i}{\partial \theta_j}\right)\right)\mathbf{b}\right)^2 \leqslant \left\{\sum_{r=1}^{N} p_r(\boldsymbol{\theta})\,\mathrm{Tr}\,\left(\sum_{j=1}^{d} b_j M_{rj}(\boldsymbol{\theta})\right)\right.$$

$$\left. \rho_r(\boldsymbol{\theta})\left(\sum_{j=1}^{d} b_j M_{rj}(\boldsymbol{\theta})\right)^{\dagger}\right\}\mathbf{a}'\,\mathrm{Cov}(X_1, \ldots, X_m|\boldsymbol{\theta})\,\mathbf{a} \qquad (3.14)$$

Let

$$\Psi_r(\boldsymbol{\theta}) = \left(\left(\mathrm{Re}\,\mathrm{Tr}\,\rho_r(\boldsymbol{\theta})M_{ri}(\boldsymbol{\theta})^{\dagger}M_{rj}(\boldsymbol{\theta})\right)\right), \quad i,j \in \{1, 2, \ldots, d\},$$

$$\Psi(\boldsymbol{\theta}) = \sum_{r=1}^{N} p_r(\boldsymbol{\theta})\Psi_r(\boldsymbol{\theta}).$$

Then the validity of (3.14) for all a_i, b_j, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant d$ implies

$$\mathrm{Cov}(X_1, X_2, \ldots, X_m|\boldsymbol{\theta}) \geqslant \left(\left(\frac{\partial f_i}{\partial \theta_j}\right)\right)\Psi^{-}(\boldsymbol{\theta})\left(\left(\frac{\partial f_i}{\partial \theta_j}\right)\right)',$$

the super index - in Ψ indicating its generalized inverse.

4. Estimators based on generalized measurements

As in Section 3 we consider a parametric family $\{\rho(\theta), \theta \in \Gamma\}$ of states of a finite level quantum system in a Hilbert space $\mathcal{H}$ and a real-valued parametric function f on Γ. In order to estimate f we now look at a generalized measurement $\mathcal{L} = (S, L)$ as described in Definition 2.1. Choose a real-valued function φ on S and if the outcome of $\mathcal{L}$ is s then evaluate $\varphi(s)$ and treat it as an estimate of $f(\theta)$. We say that $(\mathcal{L}, \varphi)$ is an *unbiased estimator* of f if

$$\sum_{s \in S} \varphi(s)\, \mathrm{Tr}\,\rho(\theta)L(s)^{\dagger}L(s) = f(\theta) \quad \forall\, \theta \in \Gamma. \qquad (4.1)$$

Indeed, it may be recalled from Section 2 that $\mathrm{Tr}\,\rho(\theta)L(s)^{\dagger}L(s)$ is the probability of the outcome s if the unknown parameter is θ. Then the variance of $(\mathcal{L}, \varphi)$ is given by

$$\mathrm{Var}(\mathcal{L}, \varphi|\theta) = \sum_{s \in S} \varphi(s)^2\,\mathrm{Tr}\,\rho(\theta)L(s)^{\dagger}L(s) - f(\theta)^2. \qquad (4.2)$$

If we write

$$X = \sum_{s \in S} \varphi(s)L(s)^{\dagger}L(s) \qquad (4.3)$$

Then X is an observable and (4.1) shows that X is an unbiased estimator of f whenever $(\mathcal{L}, \varphi)$ is an unbiased estimator of f. However, $\text{Var}(X|\theta)$ need not be the same as $\text{Var}(\mathcal{L}, \varphi|\theta)$.

In (4.1) put $T(s) = L(s)^\dagger L(s), s \in S$. Then $T(s) \geqslant 0$ and by Definition 2.1, $\sum_{s \in S} T(s) = I$. In other words $\{T(s), s \in S\}$ is a positive operator-valued distribution on S with total operator mass I. By a well-known theorem of Naimark [10], [16] one can imbed the Hilbert space $\mathcal{H}$ isometrically in a larger Hilbert space $\widehat{\mathcal{H}} = \mathcal{H} \otimes \mathcal{K}$ with $\dim \mathcal{K} < \infty$ and construct mutually orthogonal projection operators on $\widehat{\mathcal{H}}$ with the block operator form

$$E(s) = \left[\begin{array}{c|c} T(s) & M(s) \\ \hline M(s)^\dagger & N(s) \end{array}\right], \quad s \in S \tag{4.4}$$

satisfying the following:

(i) $\sum_{s \in S} E(s) = I$,

(ii) $\left\{ E(s) \begin{bmatrix} u \\ 0 \end{bmatrix}, \; s \in S, \; u \in \mathcal{H} \right\}$ spans $\widehat{\mathcal{H}}$.

Such a dilation of $T(\cdot)$ in $\mathcal{H}$ to $E(\cdot)$ in $\widehat{\mathcal{H}}$ is unique upto a natural unitary isomorphism.

Now we go back to the unbiased estimator $(\mathcal{L}, \varphi)$ of f described in (4.1). Put

$$\widehat{\rho}(\theta) \;=\; \left[\begin{array}{c|c} \rho(\theta) & 0 \\ \hline 0 & 0 \end{array}\right],$$

$$\widehat{X} \;=\; \sum_{s \in S} \varphi(s) \, E(s).$$

Then $\{\widehat{\rho}(\theta), \theta \in \Gamma\}$ is a parametric family of states in $\widehat{\mathcal{H}}$, $\widehat{X}$ is an observable in $\widehat{\mathcal{H}}$ and equations (4.1) and (4.4) imply that $\text{Tr } \widehat{\rho}(\theta)\widehat{X} = f(\theta)$. Furthermore

$$\begin{aligned} \text{Var}(\widehat{X}|\theta) \;&=\; \text{Tr } \widehat{\rho}(\theta)(\widehat{X} - f(\theta))^2 \\ &=\; \sum_{s \in S} \varphi(s)^2 \, \text{Tr } \rho(\theta)T(s) - f(\theta)^2 \\ &=\; \text{Var}(\mathcal{L}, \varphi|\theta). \end{aligned}$$

Thus $\widehat{X}$ is an unbiased estimator of f with respect to $\{\widehat{\rho}(\theta), \theta \in \Gamma\}$ with the same variance as the unbiased estimator $(\mathcal{L}, \varphi)$ based on generalized measurement for the original family of states.

If F is a Fisher map for $\{\rho(\theta), \theta \in \Gamma\}$ then $\widehat{F}$ defined by

$$\widehat{F}(\theta) = \left[\begin{array}{c|c} F(\theta) & 0 \\ \hline 0 & 0 \end{array}\right], \quad \theta \in \Gamma$$

is a Fisher map for $\{\widehat{\rho}(\theta), \theta \in \Gamma\}$ in $\widehat{\mathcal{H}}$. If $\widehat{\mathcal{I}}$ is the Fisher information form for $\{\widehat{\rho}(\theta), \theta \in \Gamma\}$ we have

$$\widehat{\mathcal{I}}(\widehat{F}_1, \widehat{F}_2)(\theta) = \mathcal{I}(F_1, F_2)(\theta).$$

Thus from Theorem 3.1 we conclude the following theorem.

Theorem 4.1. *Let $\{\rho(\theta), \theta \in \Gamma\}$ be a parametric family of states of a finite level quantum system in a Hilbert space $\mathcal{H}$ and let $(\mathcal{L}, \varphi)$ be any unbiased estimator of a real-valued parametric function f based on a generalized measurement $\mathcal{L}$ and a real scalar function φ on the set of values of the measurement. Suppose $F_j, 1 \leqslant j \leqslant n$ are Fisher maps for $\{\rho(\theta), \theta \in \Gamma\}$. Then*

$$\mathrm{Var}\left((\mathcal{L}, \varphi)|\theta\right) \geqslant (\lambda(f, F_1), \lambda(f, F_2), \ldots, \lambda(f, F_n))$$
$$\mathcal{I}_n^-\left(\lambda(f, F_1), \lambda(f, F_2), \ldots, \lambda(f, F_n)\right)'(\theta)$$

where λ is the CRB tensor and $\mathcal{I}_n^-$ is the generalized inverse of the information matrix

$$\mathcal{I}_n = ((\mathcal{I}(F_i, F_j))), \quad i, j \in \{1, 2, \ldots, n\}.$$

Proof. Immediate. ∎

We shall briefly consider the case of estimating many parametric functions $f_i(\theta)$, $1 \leqslant i \leqslant m$. In order to estimate them it appears that several generalized measurements are to be made. Such measurements have to be made in succession. As directed in Section 2 we may treat them all as a single compound generalized measurement $\mathcal{L} = (L, S)$. Let $(\mathcal{L}, \varphi_i)$ be an unbiased estimator of f_i for each i. Thus the measurement $\mathcal{L}$ is carried out and if the outcome is $s \in S$ then $\varphi_i(s)$ is the estimate of $f_i(\theta)$. The probability for the outcome s is $\mathrm{Tr}\,\rho(\theta)L(s)^\dagger L(s)$. Thus the covariance matrix of the different estimators is given by

$$\mathrm{Cov}\,(\mathcal{L}, \varphi_1, \varphi_2, \ldots, \varphi_m|\theta)$$
$$= \left(\left(\mathrm{Tr}\,\rho(\theta)\sum_{s\in S}\varphi_i(s)\varphi_j(s)L(s)^\dagger L(s) - f_i(\theta)f_j(\theta)\right)\right),$$
$$i.j \in \{1, 2, \ldots, m\}. \tag{4.5}$$

As in the discussion preceding Theorem 4.1 we can construct the Naimark dilation $\{E(s), s \in S\}$ for the positive operator-valued distribution $\{L(s)^\dagger L(s), s \in S\}$ in an enlarged Hilbert space and view the covariance matrix (4.5) as

$$\mathrm{Cov}\left(\widehat{X}_1, \widehat{X}_2, \ldots, \widehat{X}_m|\theta\right)$$

for the observables $\widehat{X}_i, = \sum_s \varphi_i(s)E(s)$ with respect to the states $\widehat{\rho}(\theta)$. This at once leads us to the CRB matrix inequality

$$\mathrm{Cov}\,(\mathcal{L}, \varphi_1, \varphi_2, \ldots, \varphi_m|\theta) \geqslant ((\lambda(f_i, F_j)))\left((\mathcal{I}_n^-(F_p, F_q))\right)$$
$$((\lambda(f_i, F_j)))'(\theta),\ 1 \leqslant i \leqslant m;\ j, p, q \in \{1, 2, \ldots, n\}.$$

for any set $\{F_j, 1 \leqslant j \leqslant n\}$ of Fisher maps, λ being the CRB tensor, $((\mathcal{I}_n(F_i, F_j)))$ the Fisher information matrix with respect to $\{F_j, 1 \leqslant j \leqslant n\}$ and the super index - denoting generalized inverse.

References

[1] E. W. Barankin, *Locally best unbiased estimators,* Ann. Math. Stat., 20, 477-501 (1949).

[2] A. Bhattacharya, *On some analogues of the amount of information and their use in statistical estimation,* Sankhyā, 8, 1-14, 201-208, 315-328 (1946, 1947, 1948).

[3] S. L. Braunstein, C. M. Caves, *Statistical distance and the geometry of quantum states,* Phys. Rev. Lett., 72, 3439-3443 (1994).

[4] H. Cramér, *Mathematical Methods of Statistics,* (Princeton University Press, Princeton, 1946).

[5] A. Dembo, T. M. Cover, J. A. Thomas, *Information theoretic inequalities,* IEEE Trans. Inform. Theory, 37, 1501-1518 (1991).

[6] C. W. Gardiner, *Quantum Noise,* (Springer Verlag, Berlin, 1991).

[7] M. Hayashi, *Quantum Information,* (Springer Verlag, Berlin, 2006).

[8] W. Heisenberg, *Uber den auschaulischen inhalt der quantentheretischen kinematik und mechanik,* Z. Phys., 43, 172-198 (1927).

[9] C. W. Helström, *Quantum Detection and Estimation Theory,* (Mathematics in Science and Engineering 123, Academic Press, New York, 1976).

[10] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory,* (North-Holland, Amsterdam, 1982); originally published in Russian (1980).

[11] M. Krishna and K. R. Parthasarathy, *An entropic uncertainty principle for quantum measurements,* Sankhyā, Ser.A, 64, 842-851 (2002).

[12] H. Maassen and J. B. M. Uffink, *Generalized entropic uncertainty relations,* Phys. Rev. Lett., 60, 1103-1106 (1988).

[13] J. Mehra, *Niels Bohr's discussions with Albert Einstein, Werner Heisenberg, and Erwin Schrödinger : the origins of the principles of uncertainty and complementarity,* in J. Mehra, The Golden Age of Theoretical Physics, Vol.2, 872-911, (World Scientific, Singapore, 2001).

[14] M. A. Nielsen and I. L. chuang, *Quantum Computation and Quantum Information,* (Cambridge University Press, Cambridge, 2000).

[15] K. R. Parthasarathy, *An Introduction to Quantum Stochastic Calculus,* (Birkhauser Verlag, Basel, 1992).

[16] K. R. Parthasarathy, *Quantum probability and strong quantum Markov processes,* in Quantum Probability Communications (Eds. R. L. Hudson, J. M. Lindsay) Vol.XII, 59-138 (World Scientific, Singapore, 2003).

[17] K. R. Parthasarathy, *Coding Theorems of Classical and Quantum Information Theory,* (Texts and Readings in Mathematics No.45, Hindustan Book Agency, New Delhi, 2007)

[18] C. R. Rao, *Information and accuracy attainable in the estimation of statistical parameters,* Bulletin of Calcutta Math. Soc., 37, 81-91 (1945).

[19] C. R. Rao, *Statistics as a last resort,* in Glimpses of India's statistical Heritage (Eds. J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy) 153-213 (Wiley Eastern, New Delhi, 1992).

[20] C. E. Shannon, *A mathematical theory of communication,* Bell System Technical Journal, 27, 379-423, 623-656 (1948).

[21] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part-I* (John Wiley and Sons, New York, 1968).

Indian Statistical Institute, Delhi Centre, 7, S. J. S. Sansanwal Marg, New Delhi - 110 016, India, e-mail: krp@isid.ac.in