Geometry of Quantum Statistical Inference

Dorje C. Brody[1,*] and Lane P. Hughston[2,†]

[1]*Blackett Laboratory, Imperial College, South Kensington, London SW7 2BZ, United Kingdom*
[2]*Merrill Lynch International, 25 Ropemaker Street, London EC2Y 9LY, United Kingdom*
and King's College London, The Strand, London WC2R 2LS, United Kingdom

An efficient geometric formulation of the problem of parameter estimation is developed, based on Hilbert space geometry. This theory, which allows for a transparent transition between classical and quantum statistical inference, is then applied to the analysis of exponential families of distributions (of relevance to statistical mechanics) and quantum mechanical evolutions. The extension to quantum theory is achieved by the introduction of a complex structure on the given real Hilbert space. We find a set of higher order corrections to the parameter estimation variance lower bound, which are potentially important in quantum mechanics, where these corrections appear as modifications to Heisenberg uncertainty relations for the determination of the parameter. [S0031-9007(96)01153-2]

A number of investigations [1] have shown that a useful approach to the study of statistical inference is to regard a parametric statistical model as a differentiable manifold equipped with a metric structure. This idea has also been applied to the study of quantum systems [2]. However, the mathematical languages used by statisticians and physicists differ, and as a consequence a clear picture of the geometrical structure of the space of probability distributions and its relation to quantum physics have not yet been fully revealed.

For example, in parametric statistics it is often convenient to deal with the log-likelihood function $l = \ln p$ instead of the density function p itself. On the other hand, in quantum mechanics we often work with a wave function, which, loosely speaking, can be thought of as a square root $\sqrt{p}$ of the probability density function. Thus, if we formulate a theory of statistical inference based upon the square-root likelihood function, the transition from classical to quantum probability should become more apparent. In particular, in the case of the square-root likelihood function, the associated natural Hilbert space norm induces a spherical geometry; that is, the "sum" $\sum (\sqrt{p})^2$ is equal to unity [3]. If the density function is parametrized by a set of parameters θ^i $(i = 1, \ldots, r)$ then for each value of θ^i we have a corresponding point on the unit sphere S in a real Hilbert space $\mathcal{H}$. Then, by choosing a suitable basis in $\mathcal{H}$ we can associate a unit vector $\xi^a(\theta^i)$ with this point. In this way, a statistical model can be characterized by a submanifold $\mathcal{M}$ in S, and the problems of statistical inference can be framed in terms of the geometry of $\mathcal{M}$ in S.

In this Letter first we reformulate some aspects of classical parametric inference, using a geometric framework. Special attention is drawn to families of exponential distributions, due to their importance in various fields in physics, especially statistical mechanics. Our main concern is the estimation problem, and especially the vari-

ance lower bound for an estimator. As is well known, the exponential distributions saturate this bound; however, for other distributions the variance exceeds the lower bound. We are thus led to establish higher-order corrections, leading to what might appropriately be called generalized Bhattacharyya bounds, which have a natural geometrical characterization. These general statistical results are then specialized to the case of quantum mechanics, by means of the introduction of a complex structure on the underlying real Hilbert space. We note that the Schrödinger evolution does not generate an exponential family of distributions in the parameter θ. This result is remarkable in demonstrating that in a problem of quantum estimation, the Heisenberg-type lower bound cannot be achieved. Thus, our results on higher-order corrections become important.

First, consider a real Hilbert space $\mathcal{H}$, equipped with a symmetric inner product denoted g_{ab}. Our idea is to represent various standard statistical operations in terms of the geometry of this space. Quantum mechanical notions will be brought in only at a later stage, with the introduction of a complex structure on $\mathcal{H}$. For convenience, we adopt a standard "abstract index" notation for Hilbert space operations [4], and write ξ^a for a typical vector in $\mathcal{H}$. Now suppose we consider the space of all probability density functions on a given configuration space. By taking their square roots we can map each such density function to a point on the unit sphere S in $\mathcal{H}$, given by $g_{ab}\xi^a\xi^b = 1$. A typical "observable" (random variable) in $\mathcal{H}$ is then represented by a symmetric bilinear form X_{ab}, with expectation given by $X_{ab}\xi^a\xi^b$ in the state ξ^a. In terms of the conventional statistical notation, one can associate ξ^a with the square-root density $p(x)^{1/2}$, X_{ab} with $x\delta(x - y)$, and hence $X_{ab}\xi^a\xi^b$ with the integral $\int_x \int_y x\delta(x - y)p(x)^{1/2}p(y)^{1/2}\,dx\,dy$. Thus, for example, $X_{ab}X^b_c\xi^a\xi^c$ is the expectation of the square of the random variable X_{ab}, and for the variance of X_{ab}

in the state ξ^a we have $\tilde{X}_{ab}\tilde{X}_c^b\xi^a\xi^c$ where $\tilde{X}_{ab} = \Delta X = X_{ab} - g_{ab}X_{cd}\xi^c\xi^d$ represents the deviation of X_{ab} from its mean in the state ξ^a.

Conversely, given the operator X_{ab} and the state ξ^a, the density function $p(x)$ can be recovered by taking the Fourier transform of the characteristic function $\phi(\lambda) = \xi^a\xi_b\exp[i\lambda X_a^b]$, whence

$$p(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\xi^a\xi_b\exp[i\lambda(X_a^b - x\delta_a^b)]\,d\lambda\,. \quad (1)$$

Alternatively, we can think of $p(x)$ as the expectation $\Delta_{ab}\xi^a\xi^b$ of the projection operator Δ_{ab} associated with the measurement outcome x, defined by

$$\Delta_{ab}(X, x) = \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp[i\lambda(X_a^c - x\delta_a^c)]\,d\lambda\right)g_{bc}\,.$$

Now consider a submanifold $\mathcal{M}$ of the unit sphere S in $\mathcal{H}$, given parametrically by $\xi^a(\theta^i)$, where θ^i are local coordinates. Then the Fisher information $\mathcal{G}_{ij}$ on $\mathcal{M}$, induced by g_{ab}, is given by the Riemannian metric

$$\mathcal{G}_{ij} = 4g_{ab}\partial_i\xi^a\partial_j\xi^b, \quad (2)$$

where $\partial_i = \partial/\partial\theta^i$. This can be seen by noticing that the squared distance between the end points of two vectors ξ^a and η^a in $\mathcal{H}$ is given by $g_{ab}(\xi^a - \eta^a)(\xi^b - \eta^b)$. If both end points lie on the submanifold $\mathcal{M}$, and η^a is obtained by infinitesimally displacing ξ^a in $\mathcal{M}$, so $\eta^a = \xi^a + \partial_i\xi^a d\theta^i$, then it follows that the separation ds between the two points on $\mathcal{M}$ is given by $ds^2 = \frac{1}{4}\mathcal{G}_{ij}d\theta^i d\theta^j$. The factor of $\frac{1}{4}$ arises in connection with the conventional definition of the Fisher information in terms of the log-likelihood function $l(x\,|\,\theta) = \ln p(x\,|\,\theta)$, given by $\mathcal{G}_{ij} = \int_x p(x\,|\,\theta)\partial_i l(x\,|\,\theta)\partial_j l(x\,|\,\theta)\,dx$. The significance of the metric $\mathcal{G}_{ij}$ is that it enables us to study the geometry of the parameter space $\mathcal{M}$. We note, for example, that the components of the standard metric connection arising from $\mathcal{G}$ are $\Gamma_{jk}^i = -\frac{1}{2}\mathcal{G}^{il}\partial_l\xi^a\partial_j\partial_k\xi_a$, from which the Riemann tensor can be calculated. Applications of this general "information geometry" to statistical physics can be found, e.g., in [5].

Suppose we are told the result of the measurement of an observable X_{ab}. We are interested in a situation where we have a one-parameter family $\xi^a(\theta)$ of possible states characterizing the distribution of the outcome of the measurement. In the case of a one-parameter family of distributions, the Fisher information is $\mathcal{G} = 4g_{ab}\dot{\xi}^a\dot{\xi}^b$, where the dot denotes $\partial/\partial\theta$. The parameter θ determines the unknown state of nature, and we wish to estimate θ with the given data. For such an estimation problem a lower bound can be established for the variance with which the estimate deviates from the true parameter value. Comparing the variance of our estimate to the lower bound, we can then enquire to what extent the estimator is optimal. A simple geometrical derivation of this bound is as follows. Given a curve $\xi^a(\theta)$ in S, we say that a random variable T_{ab} is an *unbiased estimator* for an

unknown function $\tau(\theta)$ if

$$T_{ab}\xi^a(\theta)\xi^b(\theta) = \tau(\theta)\,. \quad (3)$$

For convenience, we define a mean-adjusted estimator $\tilde{T}_{ab} \equiv T_{ab} - \tau g_{ab}$ so $\tilde{T}_{ab}\xi^a\xi^b = 0$. The variance of T is then given by $\mathrm{Var}_\xi[T] = \tilde{T}_{ab}\tilde{T}_c^b\xi^a\xi^c$, which is positive since $\mathrm{Var}_\xi[T] = g_{ab}\eta^a\eta^b$, where $\eta_b = \tilde{T}_{ab}\xi^a$. Given $T_{ab}\xi^a\xi^b = \tau$, we have $2\tilde{T}_{ab}\xi^a\dot{\xi}^b = \dot{\tau}$, and therefore $(\eta_b\dot{\xi}^b)^2 = \dot{\tau}^2/4$. By use of the Cauchy-Schwartz inequality $(\eta^a\eta_a)(\dot{\xi}^a\dot{\xi}_a) \geq (\eta_a\dot{\xi}^a)^2$, we thus obtain the following expression of the *Cramer-Rao* (CR) *inequality:*

$$\mathrm{Var}_\xi[T] \geq \frac{\dot{\tau}^2}{4\dot{\xi}^a\dot{\xi}_a}\,. \quad (4)$$

It is clear that the CR lower bound is attained only if $\dot{\xi}^a = c\eta^a$ for some constant c, which we set to $1/2$ without loss of generality. Thus, for any curve $\xi^a(\theta)$ achieving the lower bound, we obtain the differential equation

$$\dot{\xi}^a = \frac{1}{2}\tilde{T}_b^a\xi^b, \quad (5)$$

the solution to which is given by an *exponential family* of distributions, namely,

$$\xi^a(\theta) = \frac{\exp[\frac{1}{2}\theta T_b^a]q^b}{\sqrt{\exp[\theta T_b^a]q^b q_a}}\,, \quad (6)$$

where $q^a = \xi^a(0)$ is a prescribed initial distribution. The density function for the exponential family is then given by $p(x\,|\,\theta) = q(x)\exp[x\theta - W(\theta)]$, where $q(x) = \Delta_{ab}q^a q^b$ is the initial ($\theta = 0$) "background" density, and the "free energy" $W(\theta)$ is defined by

$$W(\theta) = \ln\int_{-\infty}^{\infty}e^{x\theta}q(x)\,dx = \ln(\exp[\theta T_a^b]q^a q_b)\,.$$

The mean of x is given by $\dot{W}$ and the variance by $\ddot{W}$.

Expression (6) leads us to an interesting geometrical characterization of the exponential family. We have the unit sphere S in $\mathcal{H}$, with the standard spherical metric geometry induced on it by g_{ab}. Let $\tau(\xi^a) = T_{ab}\xi^a\xi^b/\xi^c\xi_c$ be a quadratic form defined on S. Then S is foliated by hypersurfaces of constant τ. The canonical exponential family of distributions $\xi^a(\theta)$, with initial distribution q^a, is given by the unique curve through the point q^a that is everywhere orthogonal to the family of foliating τ surfaces as indicated in Fig. 1. The variance $\mathrm{Var}_\xi[T]$ at the point ξ^a is one quarter of the squared magnitude of the gradient vector $\nabla_a\tau$, which is normal to the constant τ surface at the point ξ^a. The Fisher information, on the other hand, is 4 times the squared magnitude of the tangent vector to the curve at ξ^a. Since the inner product of the tangent vector $\dot{\xi}^a$ and the normal vector $\nabla_a\tau$ is the derivative $\dot{\tau}$, it follows that $\mathrm{Var}_\xi[T] \geq \dot{\tau}^2/\mathcal{G}$, the CR inequality.

We have observed in the foregoing that the exponential family is the only class of curves achieving the variance
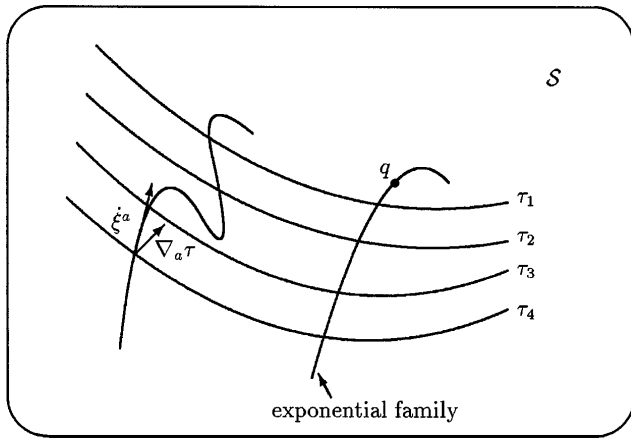
FIG. 1. Exponential families: The unit sphere S in $\mathcal{H}$ is foliated by a family of constant τ surfaces. For an exponential family, the curve generated from any given initial point q cuts the τ surfaces transversely. For other families, the tangent vector $\dot{\xi}^a$ and the normal vector $\nabla_a \tau$ are not parallel.

lower bound, providing we choose the right function of the parameter to estimate. For other families of distributions, the variance necessarily exceeds the lower bound. Hence to obtain a sharper bound we now consider the possibility of establishing higher-order corrections to the CR lower bound. Our approach closely follows that of Bhattacharyya [6]. Let us write $\xi_a^{(r)} = \partial^r \xi_a / \partial \theta^r$, the rth derivative of ξ_a with respect to θ, and write $\hat{\xi}_a^{(r)}$ for the projection of $\xi_a^{(r)}$ orthogonal to ξ_a and to all the lower order derivatives, so $\hat{\xi}_a^{(r)} \xi^a = 0$ and $\hat{\xi}_a^{(r)} \xi^{(s)a} = 0$ for $r \neq s$. Then, if T_{ab} is an estimator for the unknown function $\tau(\theta)$, so is $R_{ab} = T_{ab} + \sum_r \lambda_r \xi_{(a} \hat{\xi}_{b)}^{(r)}$ for arbitrary constants λ_r, since $R_{ab} \xi^a \xi^b = T_{ab} \xi^a \xi^b$. We only consider values of r such that $\hat{\xi}_a^{(r)} \neq 0$. A straightforward calculation leads us to the values of λ_r minimizing the variance of R, and we obtain $\min(\mathrm{Var}_\xi[R]) = \mathrm{Var}_\xi[T] - \sum_r (T_{ab} \xi^a \hat{\xi}^{(r)b})^2 / (g_{ab} \hat{\xi}^{(r)a} \hat{\xi}^{(r)b})$. Since $\mathrm{Var}_\xi[R]$ is non-negative, we deduce the following generalized *Bhattacharyya bounds* for the variance of the estimator:

$$\mathrm{Var}_\xi[T] \geq \sum_r \frac{(T_{ab} \xi^a \hat{\xi}^{(r)b})^2}{g_{ab} \hat{\xi}^{(r)a} \hat{\xi}^{(r)b}} \, . \qquad (7)$$

Clearly for $r = 1$ we recover the CR inequality. Note though that unlike the "classical" Bhattacharyya bounds, the generalized bounds given by (7) normally depend upon features of the estimator. In our applications to quantum mechanics later, however, we shall indicate an important instance of a higher-order bound that is independent of the specific choice of estimator.

As an example, we consider now the second-order correction to the CR inequality, with $r = 1, 2$. After some algebra we obtain

$$\mathrm{Var}_\xi[T] \geq \frac{(\dot{\xi}^a \nabla_a \tau)^2}{4 \dot{\xi}^a \dot{\xi}_a} + \frac{(A^a \nabla_a \tau)^2}{4 A^a A_a} \, , \qquad (8)$$

where $A_a \equiv \hat{\xi}_a^{(2)} = \ddot{\xi}_a - \dot{\xi}_a (\dot{\xi}^b \ddot{\xi}_b) / \dot{\xi}^c \dot{\xi}_c - \xi_a (\xi^b \ddot{\xi}_b)$ is

the "acceleration" vector. Note that $A^a A_a = \mathcal{G}^2 \gamma^2 / 16$, where γ^2 is the *intrinsic curvature* of the curve $\xi^a(\theta)$ in S. Since the second term in (8) is always non-negative, while the first term is just the CR lower bound, this is clearly an improvement. Furthermore, we immediately gain a geometric interpretation. As mentioned above, we have $\mathrm{Var}_\xi[T] = \frac{1}{4} g^{ab} \nabla_a \tau \nabla_b \tau$, where $\tau = T_{ab} \xi^a \xi^b / \xi^c \xi_c$. The improved CR inequality (8) states that the length of the vector $\nabla_a \tau$ is greater than the length of its orthogonal projection onto the plane spanned by velocity and acceleration vectors $\dot{\xi}^a$ and A^a, as illustrated in Fig. 2.

For the exponential families given by (6), the curvature can be calculated as $\gamma^2 = \langle \tilde{T}^4 \rangle / \langle \tilde{T}^2 \rangle^2 - \langle \tilde{T}^3 \rangle^2 / \langle \tilde{T}^2 \rangle^3 - 1$, where $\tilde{T} = \Delta T$ as defined above. Here, the first term on the right-hand side is the *kurtosis* (measure of sharpness) of the distribution, while the second term is the *skewness* (measure of asymmetry).

We now wish to extend the estimation problem to the quantum mechanical regime. In the preceding analysis, we have formulated parametric statistics in terms of intrinsic Hilbert space geometry. In order to study quantum mechanical systems based on a *real* Hilbert space $\mathcal{H}$, we need to introduce a *complex structure* on $\mathcal{H}$ compatible with the given Hilbert space geometry. The complex structure is given by a tensor J_b^a satisfying $J_b^a J_c^b = -\delta_c^a$. A symmetric operator X_{ab} is *Hermitian* if $X_{ab} J_c^a J_d^b = X_{cd}$. We require that the complex structure be compatible with the Hilbert space structure by insisting that the metric g_{ab} is Hermitian. Then, as a consequence, in the conventional Dirac notation we find that if ξ^a and η^a are two real Hilbert space vectors, their quantum mechanical product is given by $\langle \eta | \xi \rangle = \frac{1}{2} \eta^a (g_{ab} - i \Omega_{ab}) \xi^b$, where the symplectic form $\Omega_{ab} \equiv g_{ac} J_b^c$ is automatically antisymmetric and invertible. The quantum mechanical norm agrees with the real Hilbert space norm with $\langle \xi | \xi \rangle = \frac{1}{2} g_{ab} \xi^a \xi^b$. Note that a real Hilbert space element ξ^a can be decomposed into positive and negative frequency parts, with respect to the given complex structure, by writing $\xi^a = \xi_+^a + \xi_-^a$, where $\xi_\pm^a \equiv \frac{1}{2} (\xi^a \mp i J_b^a \xi^b)$. Thus we can write $\langle \eta | \xi \rangle = \eta_-^a g_{ab} \xi_+^b$.
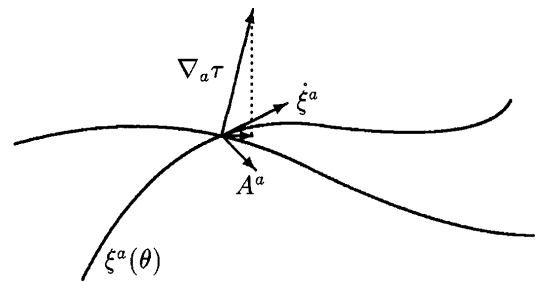

FIG. 2. Variance bounds: The gradient vector $\nabla_a \tau$ can be projected orthogonally into the plane spanned by $\dot{\xi}^a$ and A^a. The length of this projected vector is necessarily shorter than $\nabla_a \tau$.

See, e.g., [7] for further details on geometry of quantum states.

The Schrödinger equation can also be represented neatly in purely real terms. Suppose the Hamiltonian is represented by the quadratic form H_{ab}, assumed Hermitian; then the Schrödinger equation can be written as

$$\dot{\xi}^a = J_b^a \tilde{H}_c^b \xi^c, \tag{9}$$

where the parameter θ is now regarded as the time parameter t. Here the usual phase freedom in quantum mechanics is incorporated in the modified Hamiltonian $\tilde{H}_a^b = H_a^b + \varphi \delta_a^b$. If we take positive and negative parts of Eq. (9), and set $\varphi = 0$, we recover the conventional form of the Schrödinger equation $\dot{\xi}_\pm^a = \mp i H_b^a \xi_\pm^b$.

Now, suppose we set ξ^a so $\xi^a \xi_a = 1$ and fix the phase factor φ so as to minimize the Fisher information. Then, by the Schrödinger equation, we find $\varphi = -H_{ab}\xi^a\xi^b$, and thus for the Fisher information we obtain $G = 4\langle(\Delta H)^2\rangle$. Therefore, if we let T_{ab} be an unbiased estimator for the time parameter [8], with $T_{ab}\xi^a\xi^b = t$, then from Eq. (4) we find

$$\langle \tilde{T}^2 \rangle \langle \tilde{H}^2 \rangle \geq \tfrac{1}{4}. \tag{10}$$

This is formally the same as the Heisenberg uncertainty relation. However, we have observed that the condition for achieving the lower bound is for the state ξ^a to satisfy the differential equation of the form (5) above, where $\tilde{T}_{ab}$ is *symmetric*. On the other hand, the quadratic form defined by $J_b^a \tilde{H}_c^b$ in the Schrödinger Eq. (9) is *antisymmetric,* due to the Hermitian condition on the Hamiltonian. Therefore, we conclude that the Schrödinger equation cannot generate an exponential family of distributions in the t variable. A consequence of this result is that the quantum extension of the classical CR inequality does not quite provide the Heisenberg uncertainty relations. In particular, in the case of a parametric estimation for a quantum mechanical scale, the Heisenberg-type uncertainty lower bound is unattainable. For an experimentalist, the information concerning the higher-order corrections is thus indeed important in various estimation procedures.

In the case of the Schrödinger trajectories (9), we find that, in contrast to exponential families, the skewness terms drop out and the curvature is $\gamma^2 = (\langle\tilde{H}^4\rangle/\langle\tilde{H}^2\rangle^2) - 1$. Hence, the second-order correction is given by $\gamma^{-2}(T_{ab}\dot{\xi}^a\dot{\xi}^b - t\langle\tilde{H}^2\rangle)^2$, which by virtue of the Schrödinger equation (9) and the commutation relation $i[H, T] = 1$, can be reduced to an expression of the form $k\gamma^{-2}$, where k is a non-negative constant given by the covariance of the estimator T with $\tilde{H}^2$ in the initial state of the system. We seek, however, a bound on the variance that is *independent of the specific choice of estimator for t*. Remarkably, just such a term arises when we go to the third-order Bhattacharrya-style correction, which is given by $(B^a \nabla_a \tau)^2/4B^a B_a$ where $B_a \equiv \hat{\xi}_a^{(3)}$ is the component of $\dddot{\xi}_a$ orthogonal to $\xi_a, \dot{\xi}_a$, and $\ddot{\xi}_a$. Since all the odd moment terms vanish for quantum trajectories,

we find $B_a = \dddot{\xi}_a - \dot{\xi}_a(\dddot{\xi}_b\dot{\xi}^b/\dot{\xi}_c\dot{\xi}^c)ba$. Thus, after some algebra we obtain

$$\langle\tilde{T}^2\rangle\langle\tilde{H}^2\rangle \geq \frac{1}{4}\left(1 + \frac{(\langle\tilde{H}^4\rangle - 3\langle\tilde{H}^2\rangle^2)^2}{\langle\tilde{H}^6\rangle\langle\tilde{H}^2\rangle - \langle\tilde{H}^4\rangle^2}\right). \tag{11}$$

This correction is strictly non-negative. Moreover, it only depends upon the given family of probability distributions determined by $\xi^a(t)$, and is manifestly independent of the specific choice of estimator for the time parameter. The numerator in the correction is the square of the fourth cumulant of the distribution, usually denoted γ_2. The distributions for which $\gamma_2 > 0$ are called leptokurtic, and $\gamma_2 < 0$ platykurtic. If the distribution is mesokurtic ($\gamma_2 = 0$), then this correction vanishes, and an example of such a distribution is the Gaussian. For applications in quantum mechanics, however, we normally expect a distribution for H that is not Gaussian.

Our approach here has been to view statistical inference in terms of Hilbert space geometry, a view that allows us to make a firm bridge between classical and quantum statistical estimation, by introducing a complex structure. The variance lower bound obtained above is independent of the choice of estimator, and hence serves as a sharpened Heisenberg relation. This can be improved further by the incorporation of various additional higher-order terms. Although we have illustrated here the principles involved in the case of energy-time uncertainties, analogous results hold for other conjugate variables, such as position and momentum.

---

*Electronic address: d.brody@tp.ph.ic.ac.uk

†Electronic address: lane@lonnds.ml.com

[1] C. R. Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945); B. Efron, Ann. Stat. **3**, 1189 (1975); O. E. Barndorff-Nielsen, D. R. Cox, and N. Reid, Int. Stat. Rev. **54**, 83 (1986).

[2] W. K. Wootters, Phys. Rev. D **23**, 357 (1981); S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[3] A. P. Dawid, Ann. Stat. **5**, 1249 (1977); J. Burbea, Expo. Math. **4**, 347 (1986).

[4] R. Geroch, Ann. Phys. (N.Y.) **62**, 582 (1971); R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics* (Chicago University Press, Chicago, 1994).

[5] D. Brody and N. Rivier, Phys. Rev. E **51**, 1006 (1995).

[6] A. Bhattacharyya, Sankhyā **8**, 1 (1946); **8**, 201 (1947); **8**, 315 (1948).

[7] L. P. Hughston, in *Twistor Theory,* edited by S. Huggett (Marcel Dekker, Inc., New York, 1995); Proc. R. Soc. London **452**, 953 (1996); J. Anandan and Y. Aharonov, Phys. Rev. Lett. **65**, 1697 (1990); A. Ashtekar and T. A. Schilling, in *CAM-94 Physics Meeting,* edited by A. Zepeda, AIP Conf. Proc. No. 342 (AIP Press, Woodbury, NY, 1995).

[8] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland Publishing Company, Amsterdam, 1982).