Data-processing inequalities for quantum metrology

Christopher Ferrie
Center for Quantum Information and Control, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA
(Received 16 April 2014; published 21 July 2014)

We apply the classical data-processing inequality to quantum metrology to show that manipulating the classical information from a quantum measurement cannot aid in the estimation of parameters encoded in quantum states. We further derive a quantum data-processing inequality to show that coherent manipulation of quantum data also cannot improve the precision in estimation. In addition, we comment on the assumptions necessary to arrive at these inequalities and how they might be avoided, providing insights into enhancement procedures which are not provably wrong.

DOI: 10.1103/PhysRevA.90.014101 PACS number(s): 03.65.Wj, 03.67.Mn, 03.65.Ta, 06.20.−f

Parameter estimation is an integral part of physics. Quantum metrology refers to the study of the ultimate limits in the accuracy of estimates given the structure imposed by quantum theory [1,2]. Estimation at or near this limit is important for practical objectives such as improving time and frequency standards [3,4] as well as fundamental physics, such as the detection of gravitational waves [5].

Researchers have found many novel approaches to quantum metrology using, for example, multipass interferometers [6], machine learning techniques [7], and computational Bayesian statistics [8] as well as new bounds [9,10] in increasingly more general scenarios. Here we supplement these results with one of a different flavor. We provide a very general bound on the estimation accuracy in quantum metrology when noise or data processing (either classical or coherent) is present. Precisely, we give a classical and quantum *data processing inequality* which shows that estimators based on the raw data are optimal. In other words, processing the data cannot improve quantum metrology. We conclude by showing how to avoid the inequalities with more exotic procedures which can be classed into three conceptually intuitive categories: (1) processing data in a way dependent on the parameter; (2) circumventing an imposed operational restriction; or (3) modifying the dynamics which impart the parameter.

Consider a statistical model defining a likelihood function $\Pr(x|\theta;C)$. In words, there is an experimental context C whose outcomes are labeled by the random variable x and θ is an unknown parameter to be estimated. For quantum metrology, the goal is estimate a parameter which defines a quantum dynamical process:

$$\rho \mapsto \rho(\theta) := \sum_j K_j(\theta)\rho K_j^\dagger(\theta), \qquad \sum_j K_j^\dagger(\theta)K_j(\theta) = \mathbb{1}. \tag{1}$$

The statistical model is given by the structure of quantum theory and the Born rule:

$$\Pr(x|\theta;\{E_k\},\rho) = \mathrm{Tr}[\rho(\theta)E_x], \tag{2}$$

where the set $\{E_k\}$ forms a quantum measurement which defines the chosen detection strategy. In broad strokes, the goal of quantum metrology is to find the experiment context $C = (\rho,\{E_k\})$ which allows for the best accuracy in estimating θ. But how do we measure accuracy? The standard metric is *mean squared error*:

$$R(\theta,\hat{\theta};C) = \mathbb{E}_{x|\theta;C}[|\theta - \hat{\theta}(x;C)|^2], \tag{3}$$

where $\hat{\theta}$ is an *estimator*, a function which takes every possible data set to an estimate of θ. Note we have used the notation $\mathbb{E}_z[f(z)]$ to mean the expectation of the function f with respect to the distribution of z. The symbol "R" stands for "risk" and Eq. (3) denotes the risk of using the estimator $\hat{\theta}$ when the true parameter is θ.

One of the conveniences of using squared error as a measure of loss is that the risk can be lower bounded using the *Cramér-Rao bound* (CRB) [11]:

$$R(\theta,\hat{\theta};C) \geqslant I(\theta;C)^{-1}, \tag{4}$$

where $I(\theta;C)$ is the *Fisher information*:

$$I(\theta;C) = \mathbb{E}_{x|\theta;C}\left[\left(\frac{\partial}{\partial\theta}\ln \Pr(x|\theta;C)\right)^2\right]. \tag{5}$$

The CRB is a fundamental and powerful tool in statistical estimation since it bounds the performance of *every* unbiased estimator. Although the bound generally depends on the true value of the parameter, for many quantum metrology problems considered so far in the literature the Fisher information has been independent of the unknown parameter. However, this is not generally true and we must take account of the fact that θ is unknown and perhaps itself a random variable.

Suppose then that θ is a random variable with probability density $\Pr(\theta)$. Then we can remove the dependence of the risk on θ by taking a second average:

$$r(C) = \mathbb{E}_\theta[R(\theta,\hat{\theta};C)]. \tag{6}$$

The reason that r does not depend on the estimator $\hat{\theta}$ is that it is well known in statistics that the unique estimator which minimizes this quantity is [11]

$$\hat{\theta}(x;C) = \mathbb{E}_{\theta|x;C}[\theta]. \tag{7}$$

Using this, the expression for r can be simplified to

$$r(C) = \mathbb{E}_\theta[\mathbb{E}_{x|\theta;C}[|\theta - \hat{\theta}(x;C)|^2]], \tag{8}$$

$$= \mathbb{E}_{x;C}[\mathbb{E}_{\theta|x;C}[|\theta - \hat{\theta}(x;C)|^2]], \tag{9}$$

$$= \mathbb{E}_{x;C}[\mathrm{Var}_{\theta|x;C}[\theta]], \tag{10}$$

where Var denotes the variance. Note that $\Pr(\theta|x;C)$ is the *posterior* distribution using the Bayes rule:

$$\Pr(\theta|x;C) = \frac{\Pr(x|\theta;C)\Pr(\theta)}{\Pr(x;C)}. \tag{11}$$

For this reason, $r(C)$ is called the *Bayes risk*, which we have shown in Eq. (10) is the expected posterior variance, and $\hat{\theta}(x;C)$ is called the *Bayes estimator*. The Cramér-Rao bound is also generalized to the *Bayesian Cramér-Rao bound* [12]:

$$r(C) \geqslant J(C)^{-1}, \tag{12}$$

where J is the *Bayesian information*:

$$J(C) = \mathbb{E}_\theta[I(\theta;C)]. \tag{13}$$

Note that everything stated above generalizes in the expected way when $\theta \in \mathbb{R}^d$ is a vector of unknown parameters.

As stated above, quantum metrology seeks to find the experimental context which minimizes the *risk*. Or, since the bounds stated above are generally achievable (at least asymptotically), we seek to maximize the *information*. For example, the quantity

$$I_{\mathrm{Q}}(\theta) = \max_C I(\theta;C) \tag{14}$$

we call the *quantum Fisher information*. We can also define the quantity

$$J_{\mathrm{Q}} = \max_C J(C), \tag{15}$$

which we analogously call the *quantum Bayesian information*. Using these we have two *quantum* Cramér-Rao bounds:

$$R(\theta,\hat{\theta}) \geqslant I_{\mathrm{Q}}(\theta), \tag{16}$$

$$r \geqslant J_{\mathrm{Q}}. \tag{17}$$

These inequalities place the ultimate limit (called the *Heisenberg limit*) on the estimation accuracy of the unknown parameter θ. Operationally, the location of the maxima in Eqs. (14) and (15) specify the physical *experiment* which must be performed to achieve this ultimate limit.

For a fixed state ρ, the optimization over the measurement alone in Eq. (14) was introduced by Braunstein and Caves [13] and shown to be equivalent to the original definition of the quantum Fisher information given by Helstrom [14]:

$${}_{\mathrm{SLD}}I_{\mathrm{Q}}[\rho(\theta)] = \mathrm{Tr}[\rho(\theta)L(\theta)^2], \tag{18}$$

where the operator L, the *symmetric logarithmic derivative* (SLD), is implicitly defined via

$$\frac{\partial}{\partial\theta}\rho(\theta) = \frac{1}{2}[\rho(\theta)L(\theta) + L(\theta)\rho(\theta)]. \tag{19}$$

To distinguish it from the more general definition in Eq. (14), we call the definition in Eq. (18) the SLD Fisher information. As noted, the crucial difference is that the SLD Fisher information depends on ρ—that is, it is assumed that the choice of initial state is fixed. For this reason, we prefer Eq. (14) [or Eq. (15) in the Bayesian context] since it makes clear that θ is unknown and C is an experimental context, the design of the *full* experiment. This also allows us to easily restrict C when physical or practical constraints are present (such as local measurements or Gaussian states). It also makes clear that the state is part of the design, which in the general case must simultaneously be optimized [15]. On the other hand, in many cases the optimization of the measurement and preparation context can be performed separately [16], thus making the SLD Fisher information a powerful calculation tool in such cases.

Another important reason to prefer the definition of the quantum Fisher information in Eq. (14) as opposed to the symmetric logarithmic derivative version in Eq. (18) is that the latter is not generally achievable for more than a single parameter θ. In other words, to achieve the Fisher information ${}_{\mathrm{SLD}}I_{\mathrm{Q}}$ may require incompatible measurements [14]. The definition in Eq. (14) explicitly restricts the information to that achievable by valid quantum mechanical measurements.

Finally, we note that Eqs. (14) and (15) are *operational*—they tell us exactly what experimental context maximizes the information content of the measurement. With these operational definitions of information we give an operational definition of "Heisenberg limit," which is necessarily problem dependent: Given a specification of the problem, *Heisenberg limited metrology* is a realization of the experimental designs required to achieve the maximum information in either Eq. (14) or (15). This operational definition alleviates the need to resolve the recent confusion of the term [17]; the Heisenberg limit cannot be beaten because it is the limit, by definition. We can also consider restricting the allowed context $C_r \subset C$ such that the optimum cannot be achieved. For example, we could impose a restriction to laser sources and photon number constraints [16].

Another relevant restriction C_r is to that of product state inputs and outputs. In this case, the maximization of $I(\theta;C)$ or $J(C)$ over C_r is typically called the "standard quantum limit" [18]. In the special case of a restriction to independent trials, it is called the "shot noise limit." If we call such restrictions "classical," we implicitly define a *quantum resource*: those experimental contexts in $C\backslash C_r$ whose information is larger than that maximized over C_r.

Having specified the problem, we will now apply the so-called *data processing inequality* to the quantum metrology to show that postprocessing of the data can never improve the estimation accuracy. First, a definition: $\theta \to x \to y$ is called a *Markov chain* if $\Pr(y,x,\theta) = \Pr(y|x)\Pr(x|\theta)\Pr(\theta)$. Note that if y is some deterministic function (a *statistic*) of x, that is $y = f(x)$, then $\theta \to x \to f(x)$ is trivially a Markov chain. Why is this relevant to estimation? The chain $\theta \to x \to y$ can be thought of as an estimation procedure where θ generates the raw data x via the statistical model $\Pr(x|\theta)$ and then that data is *postprocessed* (in general, probabilistically) to arrive at y. The information flowing through the chain can be used to estimate θ. Next, we show that the second step, postprocessing, cannot improve the estimation accuracy.

The first data processing inequality applies to the Fisher information and is [19]

$$I_y(\theta) \leqslant I_x(\theta), \tag{20}$$

with equality if and only if $\theta \to y \to x$ is also a Markov chain [which is equivalent in the case $y = f(x)$ to f being a *sufficient statistic*]. This inequality implies the analogous

Bayesian information variant:

$$J_y \leqslant J_x. \tag{21}$$

Both inequalities state that the Fisher (respectively, Bayesian) information calculated using the distribution of processed data $\Pr(y|\theta)$ is less than that computed using the original distribution of raw data $\Pr(x|\theta)$. Then the Cramér-Rao bounds state the mean squared error of an unbiased estimator of θ is worse when postprocessing.

Let us apply this to the quantum metrology setting where the conclusion should be unsurprising. Indeed, it is quite simple to include an additional experimental context C in the classical description above. Let us start with the Fisher information version first. The data processing inequality in Eq. (20) remains unchanged when adding an additional context:

$$I_y(\theta; C) \leqslant I_x(\theta; C). \tag{22}$$

Since this holds for all C, it holds where each side individually obtains its maximum. That is,

$$\max_C I_y(\theta; C) \leqslant \max_C I_x(\theta; C). \tag{23}$$

This is the quantum Fisher information when using either the raw data x or postprocessed data y:

$$I_{y,Q}(\theta) \leqslant I_{x,Q}(\theta). \tag{24}$$

Then, the quantum Cramér-Rao bound implies that conditioning on postprocessed data cannot improve the estimation of θ. The same argument applies to Eq. (21). If we add the context C and maximize, we find

$$J_{y,Q} \leqslant J_{x,Q}. \tag{25}$$

The Bayesian Cramér-Rao bound then implies that the Bayes risk of using postprocessed data is higher.

The above results apply to the case where "data processing" refers to classical computation of classical data. Perhaps it might be the case that coherent data processing—quantum computation of quantum data—might aid in the estimation of the parameters θ. In this case, rather than the classical process $\theta \to x \to y$, we have the quantum process $\theta \to \rho(\theta) \to \mathcal{E}[\rho(\theta)]$, where $\mathcal{E}$ is a quantum operation (completely positive, trace preserving map). Next, we prove a *quantum data processing inequality* which analogously shows that coherent manipulation of data also cannot aid quantum metrology.

The result is as follows. If $\mathcal{E}$ is a quantum operation, then

$$I_{\mathcal{E},Q}(\theta) \leqslant I_Q(\theta). \tag{26}$$

The proof is remarkably simple. First note that

$$\Pr[x|\theta; \{E_k\}, \mathcal{E}(\rho)] = \mathrm{Tr}\{\mathcal{E}[\rho(\theta)]E_x\}, \tag{27}$$

$$= \mathrm{Tr}[\rho(\theta)\mathcal{E}^\dagger(E_x)], \tag{28}$$

$$= \Pr(x|\theta; \{\mathcal{E}^\dagger(E_k)\}, \rho), \tag{29}$$

where $\mathcal{E}^\dagger$ is the *dual channel*—a *Heisenberg picture* for quantum channels. Explicitly, if the map $\mathcal{E}$ has the Kraus decomposition

$$\mathcal{E}(\rho) = \sum_j K_j \rho K_j^\dagger, \tag{30}$$

then

$$\mathcal{E}^\dagger(E_k) = \sum_j K_j^\dagger E_k K_j. \tag{31}$$

In words, the act of subjecting $\rho(\theta)$ to an additional quantum channel is equivalent to subjecting the measurement to the dual channel.

Now, since it is the measurement to be optimized, either the range of $\mathcal{E}^\dagger$ contains the optimal measurement, or it does not. The channel $\mathcal{E}^\dagger$ serves only to restrict the possible measurements. That is,

$$\max_{\rho,\{\mathcal{E}^\dagger(E_k)\}} I(\theta; \rho, \{\mathcal{E}^\dagger(E_k)\}) \leqslant \max_{\rho,\{E_k\}} I(\theta; \rho, \{E_k\}). \tag{32}$$

Thus, by definition,

$$I_{\mathcal{E},Q}(\theta) \leqslant I_Q(\theta). \tag{33}$$

This is the quantum data processing inequality and it states that no coherent manipulation of the data allowed by quantum theory improves the estimation accuracy of θ.

Some comments are in order. First, we note the that the temporal order of the data processing is irrelevant. The quantum process $\theta \to \rho(\theta) \to \mathcal{E}[\rho(\theta)]$ has the channel $\mathcal{E}$ act after the parameter has been imparted. However, the conclusion remains if the process is $\rho \to \mathcal{E}(\rho) \to \mathcal{E}(\rho)(\theta)$. That is, Eq. (26) holds if $\mathcal{E}$ refers to "preprocessing" or "encoding."

Secondly, we comment on the the terminology "data processing inequality." This term is more popularly used in the context of information theory, where it applies to the *mutual information* between either y and θ or x and θ in the Markov chain $\theta \to x \to y$. If $\mathcal{I}(a; b)$ denotes the mutual information (a measure of correlations) between a and b, then the more commonly used *data processing inequality* is $\mathcal{I}(\theta; y) \geqslant \mathcal{I}(\theta; x)$ (see, for example, [20]). In words, it says the same thing as the inequality we have used here (proven in [19]): Manipulating the data cannot increase the amount of information one has about θ. This information theoretic data processing inequality is not directly applicable to estimation but is a fundamental result in information theory. As one might expect, then, it has been generalized to the quantum mechanical setting [21].

The next thing to mention is noise. Note that, in the classical setting the only assumption was that $\theta \to x \to y$ was a Markov chain. It need not be the case that y is some deterministic function of x. So, the channel $x \to y$ could also represent classical technical noise on the detector. So long as the noise is statistically independent of the unknown parameter θ given x, the data processing inequality applies. Thus, noise assisted metrology cannot be realized. Similarly, in the quantum channel setting, $\mathcal{E}$ could represent a decoherence mechanism rather than a purposefully built quantum circuit. The conclusion remains; decoherence cannot improve estimation accuracy.

The final comment is on "outs." How do we avoid this conclusion? The three most natural possibilities are as follows:

(O1) have $\mathcal{E}$ depend on θ;

(O2) arrange for $\mathcal{E}$ to circumvent an additional imposed restriction on the allowed context C; or

(O3) modify the dynamics in Eq. (1) which impose the parameter.

We will discuss these possibilities with reference to examples in the literature. Experimentally relevant examples of **O1** include the multipass interferometer of Ref. [6] and adaptive measurement techniques of Refs. [22] and [8]. In these phase estimation examples, additional channels are added (quantum data processing) which depend on the unknown parameter. In [6], each pass through the interferometer accumulates more phase difference on the state. In [22] and [8] the phase accumulation time is chosen adaptively based on (current, yet still incomplete) knowledge of the unknown parameter(s).

Nontrivial examples of **O2** are more common, which might be expected since operational restrictions are often artificially imposed by us—either through our experimental choices or what is currently technologically feasible. In this vein, entanglement producing operations can be seen as a data processing resource to increase sensitivity beyond what can be achieve by separable states [2]. In the context of high precision interferometry, current infrastructure and technology finds that a high-power input laser supplemented by squeezed states [1] is the most practical [5]. This restriction and its limits [16] can be overcome through the data processing resource of creating the absolute optimal NOON state input [23].

The NOON states, on the other hand, are difficult (at present) to create and suffer greatly from noise [24], which brings us to **O3**. In situation **O3**, we are imagining something conceptually different [25]. Here we are changing the problem itself through a modification of how the dynamics impose the unknown parameter. An example of this is a quantum error correction inspired approach—recently rediscovered—to dynamically correct errors, interleaving the imposition of the parameter with recovery operations [26,27]. The examples presented in [26,27] correct bit-flip errors which allows the parameter to be imparted via single qubit logical operations. Similar protocols can recover from arbitrary single qubit errors, but require the more difficult task of imparting the parameter through high-weight logical operations in a quantum error correcting code [28,29]. Note that the improvement is not simply due to the encoding of the state—such a preprocessing procedure would still be bounded by the data processing inequality—but the active changing of the way the dynamics impose the parameter on the encoded state. That an encoded state is optimal for this task is implicit through the maximization over input states in the definition of the quantum Fisher information.

Quantum metrology can be thought of as a purely statistical problem. Often, thinking of quantum mechanical problems classically leads to paradoxes or, in the very least, is just cumbersome—which is why concepts like the SLD quantum Fisher information exist. However, if we are careful to avoid the usual pitfalls, rephrasing quantum metrology in classical language allows us to leverage known classical results. In particular, we have applied the data processing inequalities to show that postprocessing raw data cannot lead to more precise estimates of parameters. The classical picture then allows for a simple generalization, which we have called the quantum data processing inequality, showing that coherent data processing suffers the same restriction. Finally, the classical representation of these results displays more transparently the assumptions necessary to provide this curtailment thus readily allowing us to provide operationally meaningful statements of how to avoid the inequalities. We hope these considerations shed light on the myriad of definitions of "standard quantum limit," "Heisenberg limit," and so on, and perhaps make conceptually clear why and when one can improve on standard estimation procedures.

The author thanks Carl Caves, Josh Combes, Zhang Jiang, and Chris Granade for helpful discussions. This work was supported in part by National Science Foundation Grants No. PHY-1212445 and No. PHY-1314763, by Office of Naval Research Grant No. N00014-11-1-0082, and by the Canadian Government through the NSERC PDF program.

[1] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photonics **5**, 222 (2011).
[3] T. Udem, R. Holzwarth, and T. W. Hänsch, Nature (London) **416**, 233 (2002).
[4] N. Hinkley, J. Sherman, N. Phillips, M. Schioppo, N. Lemke, K. Beloy, M. Pizzocaro, C. Oates, and A. Ludlow, Science **341**, 1215 (2013).
[5] J. Aasi, J. Abadie, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Nat. Photonics **7**, 613 (2013).
[6] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature (London) **450**, 393 (2007).
[7] A. Hentschel and B. C. Sanders, Phys. Rev. Lett. **107**, 233601 (2011).
[8] C. E. Granade, C. Ferrie, N. Wiebe, and D. Cory, New J. Phys. **14**, 103013 (2012).
[9] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, Phys. Rev. Lett. **98**, 090401 (2007).
[10] M. Tsang, H. M. Wiseman, and C. M. Caves, Phys. Rev. Lett. **106**, 090401 (2011).
[11] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer, New York, 1998).
[12] R. D. Gill and B. Y. Levit, Bernoulli **1**, 59 (1995).
[13] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).
[14] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
[15] M. Hayashi, Commun. Math. Phys. **304**, 689 (2011).
[16] M. D. Lang and C. M. Caves, Phys. Rev. Lett. **111**, 173601 (2013).
[17] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. Lett. **105**, 180402 (2010).
[18] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[19] O. Rioul, IEEE Trans. Inf. Theory **57**, 33 (2011).
[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2012).
[21] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).
[22] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Phys. Rev. A **84**, 052315 (2011).
[23] H. Lee, P. Kok, and J. P. Dowling, J. Mod. Opt. **49**, 2325 (2002).
[24] U. Dorner, R. Demkowicz-Dobrzanski, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Phys. Rev. Lett. **102**, 040403 (2009).
[25] The distinction between **O1** and **O3** is subtle: **O1** can be thought of as a subset of **O3** but not always vice versa.
[26] J. Preskill, arXiv:quant-ph/0010098.
[27] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. **112**, 150802 (2014).
[28] W. Dür, M. Skotiniotis, F. Fröwis, and B. Kraus, Phys. Rev. Lett. **112**, 080801 (2014).
[29] G. Arrad, Y. Vinkler, D. Aharonov, and A. Retzker, Phys. Rev. Lett. **112**, 150801 (2014).