Quantum estimation theory

Jader P. Santos

1 Estimation theory [1]

1.1 Introduction

Given a parameter of interest, such as a population mean μ, the objective of estimation theory is to use a sample to compute a number that represents in some sense a good guess for the true value of the parameter. The resulting number is called a point estimate or estimative. Obtaining a estimative entails calculating the value of a statistic[2] such as the sample mean $\bar{X}$ or sample standard deviation S.

When discussing general concepts and methods of inference, it is convenient to have a generic symbol for the parameter of interest. We will use the letter θ for this purpose. The objective of a estimation is to select a single number, based on sample data, that represents a sensible value for θ. The estimative of a parameter θ is obtained by selecting a suitable statistic and computing its value from a the given sample data. The selected statistic is called the estimator of θ. The symbol $\hat{\theta}$ is customarily used to denote both the estimator of θ and the estimative resulting from a given sample.

Example 1.1. A natural estimator for the population variance σ^2 is the sample variance:

$$\hat{\sigma}^2 = S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1} \tag{1}$$

Here X_i is a random variable. An alternative estimator would result from using divisor n instead of $n-1$ (i.e., the average squared deviation):

$$\hat{\sigma}^2 = S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n} \tag{2}$$

We shall see sun why Eq. (1) is a better estimator for the variance than Eq. (2).

In the best of all possible worlds, we could find an estimator $\hat{\theta}$ for which $\hat{\theta} = \theta$ always. However, $\hat{\theta}$ is a function of the sample X_i's, so it is a random variable. For some sample, $\hat{\theta}$ will yield a value larger than θ, whereas for other samples $\hat{\theta}$ will underestimate θ. If we write

$$\hat{\theta} = \theta + \text{error of estimation} \tag{3}$$

[1] Most of this text is a summary of some ideas presented in Ref. [1]

[2] A statistic is any quantity whose value can be calculated from sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result. Therefore, a statistic is a random variable and will be denoted by an uppercase letter.

then an accurate estimator would be one resulting in small estimation errors, so that estimated values will be near the true value.

1.2 Mean Squared Error

The mean squared error of an estimator $\hat{\theta}$ is[3] $E[(\hat{\theta}-\theta)^2]$. Note that by using the variance $V(Y)$,

$$V(Y) = E(Y^2) - [E(Y)]^2 \tag{4}$$

we can write

$$\begin{aligned} E[(\hat{\theta}-\theta)^2] &= V(\hat{\theta}-\theta) + [E(\hat{\theta}-\theta)]^2 \\ &= \underbrace{V(\hat{\theta})}_{\text{variance of estimator}} + \underbrace{[E(\hat{\theta})-\theta]^2}_{(\text{bias})^2} \end{aligned} \tag{5}$$

1.3 Unbiased Estimator

An estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ. If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the bias of $\hat{\theta}$.

Example 1.2. Let us turn to the problem of estimating σ^2 based on a random sample $X_1,\ldots,X_n$. First consider the estimator[a],

$$\begin{aligned} S^2 &= \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n-1} = \sum_{i=1}^{n} \frac{(X_i^2 - 2X_i\bar{X} + \bar{X}^2)}{n-1} \\ &= \frac{(\sum_{i=1}^{n} X_i^2) - 2(\sum_{i=1}^{n} X_i)\bar{X} + (\sum_{i=1}^{n} \bar{X}^2)}{n-1} \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right] \end{aligned} \tag{6}$$

Now if we calculate the average value of S^2 we have

$$\begin{aligned} E(S^2) &= \frac{1}{n-1}\left\{\sum E[X_i^2] - \frac{1}{n}E\left[\left(\sum X_i\right)^2\right]\right\} \\ &= \frac{1}{n-1}\left\{\sum(\sigma^2+\mu^2) - \frac{1}{n}\left\{V\left(\sum X_i\right) + \left[E\left(\sum X_i\right)\right]^2\right\}\right\} \\ &= \frac{1}{n-1}\left\{n\sigma^2 + n\mu^2 - \frac{1}{n}n\sigma^2 - \frac{1}{n}(n\mu)^2\right\} \\ &= \frac{1}{n-1}\{n\sigma^2 - \sigma^2\} = \sigma^2 \end{aligned} \tag{7}$$

Than we have show that the sample variance S^2 is an unbiased estimator of σ^2. The estimator that uses the divisor n can be expressed as $(n-1)S^2/n$, so

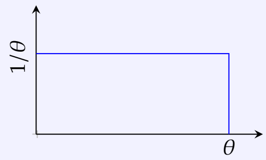
$$E\left[\frac{(n-1)S^2}{n}\right] = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2 \tag{8}$$

This estimator is therefore biased.

[a] Here $\bar{X} = \sum_{i=1}^{n} X_i/n$

[3] Here $E(X)$ mean the expected value (or the mean value) of the random variable X.

Example 1.3. Suppose that X has an uniform distribution on the interval from 0 to an unknown upper limit θ. We want to estimate θ on the basis of a random sample $X_1, X_2, \ldots, X_n$.



Since θ is the largest possible value of the entire population, consider as a first estimator the largest sample value:

$$\hat{\theta}_b = \max(X_1, \ldots, X_n). \tag{9}$$

Note we have,

$$E[\hat{\theta}_b] = \frac{n}{n+1}\theta \tag{10}$$

Thus, $\hat{\theta}_b$ is a biased estimator. It is easy to modify $\hat{\theta}_b$ to obtain an unbiased estimator of θ. Consider the estimator

$$\hat{\theta}_u = \frac{n+1}{n}\hat{\theta}_b \tag{11}$$

and now $E[\hat{\theta}_u] = \left(\frac{n+1}{n}\right)E[\hat{\theta}_b] = \theta$.

1.4 Estimator with Minimum Variance

Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators of θ that are both unbiased. Then, although the distribution of each estimator is centered at the true values of θ, the spread of the distributions about the true values may be different. Among all estimators of θ that are unbiased, the one that has minimum variance is called the **minimum variance unbiased estimator (MVUE)** of θ.

Seeking an unbiased estimator with minimum variance is the same as seeking an unbiased estimator that has minimum mean squared error.

Example 1.4. When $X_1 \ldots, X_n$ is a random sample from a uniform distribution on $[0, \theta]$, the estimator

$$\hat{\theta}_1 = \frac{n-1}{n}\max(X_1, \ldots, X_n) \tag{12}$$

is a unbiased estimator for θ, i.e. $E[\hat{\theta}_1] = \theta$. This is not the only unbiased estimator of θ. Note that $E[X_i] = \theta/2$. This implies that $E[\bar{X}] = \theta/2$, from which $E[2\bar{X}] = \theta$, i.e. the estimator $\hat{\theta}_2 = 2\bar{X}$ is unbiased for θ. It is possible to show that,

$$V(\hat{\theta}_1) = \frac{\theta^2}{n(n+2)} \quad \text{and} \quad V(\hat{\theta}_2) = \frac{\theta^2}{3n} \tag{13}$$

As long as $n > 1$, $V(\hat{\theta}_1) < V(\hat{\theta}_2)$, so $\hat{\theta}_1$ is a better estimator than $\hat{\theta}_2$. More advanced methods can be used to show that $\hat{\theta}_1$ is the MVUE of θ.

1.5 The standard error of the estimator

Besides reporting the value of a point estimate, some indication of its precision should be given. The usual measure of precision is the standard error of the estimator. The standard error of an estimator $\hat{\theta}$ is its standard deviation

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})} \tag{14}$$

1.6 Methods of point estimation

We now discuss two methods for obtaining estimators: the method of moments and the method of maximum likelihood.

1.6.1 The method of moments

The basic idea of this method is to equate certain sample characteristics, such as the mean, to the corresponding population expected values. Then solving these equations for unknown parameters values yields the estimators.

Let us start by considering $X_1, \ldots, X_n$ random sample from a probability distribution function $p(x)$. For $k = 1, 2, \ldots$, the kth population moment, or kth moment of the distribution $p(x)$, is $E(X^k)$. The kth sample moment is $(1/n)\sum_{i=1}^{n} X_i^k$. Thus the first population moment is $E(X) = \mu$ and the first sample moment is $\sum X_i/n = \bar{X}$. The second population and sample moment are $E(X^2)$ and $\sum X_i^2/n$, respectively. The population moments will be functions of any unknown parameters $\theta_1, \theta_2 \ldots$.

Now, let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with probability distribution function $p(x|\theta_1, \ldots, \theta_m)$, where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown. Then the moment estimators $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are obtained by equating the first m sample moments to the corresponding first m populations moments and solving for $\theta_1, \ldots, \theta_m$.

Example 1.5. Let $X_1, \ldots, X_n$ be a random sample from a gamma distribution

$$p(x|\alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where $\alpha > 0$ and $\beta > 0$. It is possible to show that,

$$E[X] = \alpha\beta \quad \text{and} \quad E[X^2] = \alpha\beta^2(\alpha + 1) \tag{16}$$

The moments estimators of α and β are obtained by solving

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} = \alpha\beta \quad \text{and} \quad \sum_{i=1}^{n} \frac{X_i^2}{n} = \alpha(\alpha + 1)\beta^2 \tag{17}$$

From the above result, it is possible to write the estimators

$$\hat{\alpha} = \frac{(\bar{X})^2}{\frac{1}{n}\sum X_i^2 - (\bar{X})^2} \quad \text{and} \quad \hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - (\bar{X})^2}{\bar{X}} \tag{18}$$

1.6.2 Maximum Likelihood Estimation[4]

Let $X_1, \ldots, X_n$ have joint probability distribution function

$$p(x_1, x_2, \ldots, x_n | \theta_1, \ldots, \theta_m) \tag{19}$$

where the parameters $\theta_1, \ldots, \theta_m$ have unknown values. When $x_1, \ldots, x_n$ are the observed sample values and Eq. (19) is regarded as a function of $\theta_1, \ldots, \theta_m$ it is called the **likelihood function**. The maximum likelihood estimates $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are those values of the θ_i that maximize the likelihood function, so that

$$p(x_1, x_2, \ldots, x_n | \hat{\theta}_1, \ldots, \hat{\theta}_m) \geq p(x_1, x_2, \ldots, x_n | \theta_1, \ldots, \theta_m) \text{ for all } \theta_1, \ldots, \theta_m \tag{20}$$

When the X_i are substituted in place of the x_i's, the **maximum likelihood estimators** result.

The likelihood function tells us how likely the observed sample is as a function of the possible parameter values. Maximizing the likelihood gives the parameter values for which the observed sample is most likely to have been generated, that is, the parameter values that "agree most closely" with the observed data.

Example 1.6. Suppose $X_1, \ldots, X_n$ is a random sample from an exponential distribution with parameter λ. Because of independence, the likelihood function is a product of the individual probability distribution function:

$$p(x_1, \ldots, x_n | \lambda) = (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \sum x_i} \tag{21}$$

The logarithmic of the likelihood is

$$\ln\left[p(x_1, \ldots, x_n | \lambda)\right] = n \ln \lambda - \lambda \sum x_i \tag{22}$$

Now we can make

$$\frac{d}{d\lambda}\left[\ln(p(x_1, \ldots, x_n | \lambda))\right] = 0 \implies \frac{n}{\lambda} - \sum x_i = 0 \implies \lambda = \frac{n}{\sum x_i} \tag{23}$$

Thus the maximum likelihood estimator is

$$\hat{\lambda} = \frac{1}{\bar{X}} \tag{24}$$

Example 1.7. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution. The likelihood function is

$$f(x_1, \ldots, x_n | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/(2\sigma^2)} \ldots e^{-(x_n-\mu)^2/(2\sigma^2)} \tag{25}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\sum(x_i-\mu)^2/(2\sigma^2)} \tag{26}$$

so

$$\ln\left[f(x_1, \ldots, x_n | \mu, \sigma^2)\right] = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \tag{27}$$

To find the Maximizing values of μ and σ^2, we must take the partial

[4] Introduced by R. A. Fisher in the 1920s. You can read more about in Ref. [2].

derivatives of $\ln p$ with respect to μ and σ^2 and equating them to zero,

$$\frac{\partial}{\partial\mu}\ln\left[f\right]=0 \implies \mu=\sum\frac{x_i}{n} \implies \hat{\mu}=\bar{X} \tag{28}$$

$$\frac{\partial}{\partial\sigma^2}\ln\left[f\right]=0 \implies \hat{\sigma}^2=\sum\frac{(x_i-\mu)^2}{n} \implies \hat{\sigma}^2=\sum\frac{(X_i-\bar{X})^2}{n} \tag{29}$$

The Maximum likelihood estimator of σ^2 is not the unbiased estimator, so two different principles of estimation yield two different estimators.

Under very general conditions on the joint distribution of the sample, when the sample size is large, the maximum likelihood estimator of any parameter θ is close to θ (consistency), is approximately unbiased [$E(\hat{\theta})\approx\theta$], and has variance that is nearly as small as can be achieved by any unbiased estimator. Stated another way, the maximum likelihood estimator $\hat{\theta}$ is approximately the MVUE of θ.

1.7 Information and Efficiency

Consider $p(x|\theta)$ a probability density function with unknown parameter θ. The Fisher information is intended to measure the precision in a single observation. Consider a random variable U obtained by taking the partial derivative of $\ln[p(x|\theta)]$ with respect to θ and then replacing[5] x by X:

$$U=\frac{\partial}{\partial\theta}[\ln[p(X;\theta)]] \tag{30}$$

the **Fisher information** $F(\theta)$ in a single observation from a probability density function $p(x|\theta)$ is the variance of the random variable $U=\partial_\theta\Big[\ln(p(X|\theta))\Big]$

$$\boxed{F(\theta)=V\Big[\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big]} \tag{31}$$

There is an alternative expression for $F(\theta)$ that is sometimes easier to compute:

$$\boxed{F(\theta)=E\Big[\Big(\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big)^2\Big]=\sum_x p(x|\theta)\Big[\frac{\partial}{\partial\theta}\ln(p(x|\theta))\Big]^2} \tag{32}$$

Proof. Let us start by Let us calculate the average value of U,

$$E[U]=E\Big[\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big]$$
$$E[U]=\sum_x p(x|\theta)\frac{\partial}{\partial\theta}\ln(p(x|\theta))=\sum_x\frac{\partial}{\partial\theta}p(x|\theta)=\frac{\partial}{\partial\theta}\sum_x p(x|\theta)=0 \tag{33}$$

[5] Remember that here X denotes a random variable.

Now we get,

$$F(\theta) = V\Big[\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big] = E\Big[\Big(\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big)^2\Big] - E\Big[\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big]^2 \to 0$$

$$F(\theta) = E\Big[\Big(\frac{\partial}{\partial\theta}\ln(p(X|\theta))\Big)^2\Big] \tag{34}$$

□

Example 1.8 (Fisher Information). Let X be a Bernoulli random variable, so $p(x, r) = r^x(1-r)^{1-x}$ and $x = 0, 1$. Then

$$F(r) = \sum_{x=0}^{1} p(x|r)\Big[\frac{\partial}{\partial r}\ln(p(x|r))\Big]^2 = \frac{1}{r(1-r)} \tag{35}$$

and

$$V(X) = E[X^2] - E[X]^2 = r(1-r) = 1/F(r) \tag{36}$$

which says that the information is the reciprocal of $V(X)$. It is reasonable that the information is greatest when the variance is smallest.

1.8 Information in a Random Sample

Let us assume a random sample $X_1, X_2, \ldots, X_n$ from a distribution $p(x|\theta)$. Let $p(X_1, X_2, \cdots, X_n|\theta) = p(X_1|\theta)p(X_2|\theta)\cdots p(X_n|\theta)$ be the likelihood function. The Fisher information $I_n(\theta)$ for the random sample is the variance of the function $\partial_\theta[\ln(p(X|\theta))]$. Then

$$\begin{aligned}\partial_\theta[\ln(p(X|\theta))] &= \frac{\partial}{\partial\theta}\ln p(X_1, X_2, \ldots, X_n|\theta)\\ &= \frac{\partial}{\partial\theta}\ln\big[p(X_1|\theta)p(X_2|\theta)\ldots p(X_n|\theta)\big]\\ &= \frac{\partial}{\partial\theta}\ln p(X_1|\theta) + \cdots + \frac{\partial}{\partial\theta}\ln p(X_n|\theta)\end{aligned} \tag{37}$$

Taking the variance of both sides of Eq. (37) gives the information $F_n(\theta)$ in the random sample,

$$F_n(\theta) = V\Big[\frac{\partial}{\partial\theta}\ln p(X_1, X_2, \ldots, X_n|\theta)\Big] = nV\Big[\frac{\partial}{\partial\theta}\ln p(X_1|\theta)\Big] = nF(\theta) \tag{38}$$

Therefore, the Fisher information in a random sample is just n times the information in a single observation. This should make sense intuitively, because it says that twice as many observations yield twice as much information.

1.9 The Cramér-Rao Inequality

Assume a random sample $X_1, X_2, \ldots, X_n$ from the distribution with probability distribution function $p(x|\theta)$ such that the set of possible values does not depended on θ. If the statistic $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ is an unbiased estimator for the parameter θ, then

$$\boxed{V(\hat{\theta}) \geq \frac{1}{V\left[\frac{\partial}{\partial\theta}[\ln p(X_1, \ldots, X_n|\theta)]\right]} = \frac{1}{nV\left[\frac{\partial}{\partial\theta} \ln p(X_1|\theta)\right]} = \frac{1}{nF(\theta)}} \tag{39}$$

The ration of the lower bound to the variance of $\hat{\theta}$ is the efficiency. Then $\hat{\theta}$ is said to be an efficient estimator if $\hat{\theta}$ achieves the Cramér-Rao lower bound. An efficient estimator is a minimum variance unbiased estimator (MVUE).

2 Quantum estimation[6]

Several quantities of interest in quantum information, including entanglement and purity, are nonlinear functions of the density matrix and cannot, even in principle, correspond to proper quantum observables. Any method aimed to determine the value of these quantities should resort to indirect measurements and this corresponds to a parameter estimation problem whose solution, i.e. the determination of the most precise estimator unavoidably involves an optimization procedure.

The solution of a parameter estimation problem amounts to find an estimator, i.e. a mapping $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots)$ from a set χ of measure outcomes into the set of parameters. As we saw in Sec. 1, optimal estimators in classical estimation theory are those saturating the Cramér-Rao inequality,

$$V(\hat{\theta}) \geq \frac{1}{nF(\theta)} \tag{40}$$

which establish a lower bound on the mean square error $V(\hat{\theta}) = E[(\hat{\theta}(\{X\}) - \theta)^2]$ of any estimator of the parameter θ. In Eq. (40) n is the number of measurements and $F(\theta)$ is the Fisher Information

$$F(\theta) = \int \mathrm{d}x\, p(x|\theta) \Big[\frac{\partial \ln p(x|\theta)}{\partial \theta}\Big]^2 = \int \mathrm{d}x\, \frac{1}{p(x|\theta)} \Big[\frac{\partial p(x|\theta)}{\partial \theta}\Big]^2 \tag{41}$$

where $p(x|\theta)$ denotes the conditional probability of obtaining the value x when the parameter has the value θ. For unbiased estimators, as those we will deal with, the mean square error is equal to the variance

$$V(\hat{\theta}) = E[\hat{\theta}^2] - E[\hat{\theta}]^2 \tag{42}$$

The parameter θ that we want to estimate does not, in general, correspond to a quantum observable and our aim is to estimate its values thorough the measurement of some observable. A **quantum estimator** O_θ for θ is a selfadjoint operator, which describe a quantum measurement followed by any classical data processing performed on the outcomes. In quantum mechanics, according to the Born rule we have $p(x|\theta) = \mathrm{Tr}[\Pi_x \rho_\theta]$ where $\{\Pi_x\}$ are the elements of a positive operator-value measurement (POVM)[7], and ρ_θ is the density operator parametrized by the quantity we want to estimate.

2.1 Quantum Cramér-Rao Bound

Symmetric Logarithmic Derivate (SLD)

Let us introduce the Symmetric Logarithmic Derivative L_θ as the selfadjoint operator ($L_\theta^\dagger = L_\theta$) satisfying the equation

$$\frac{L_\theta \rho_\theta + \rho_\theta L_\theta}{2} = \frac{\partial \rho_\theta}{\partial \theta} \tag{43}$$

[6]This section is a summary of the nice work from Matteo G. A. Paris [3]

[7]Note that $\int \mathrm{d}x\, \Pi_x = 1$

Note that

$$\begin{aligned}
\partial_\theta p(x|\theta) &= \partial_\theta \operatorname{Tr}\{\Pi_x \rho_\theta\} = \operatorname{Tr}\{\Pi_x \partial_\theta \rho_\theta\} \\
&= \operatorname{Tr}\left\{\Pi_x \Big(\frac{L_\theta \rho_\theta + \rho_\theta L_\theta}{2}\Big)\right\} \\
&= \frac{1}{2}\operatorname{Tr}\{\Pi_x L_\theta \rho_\theta\} + \frac{1}{2}\operatorname{Tr}\{\Pi_x \rho_\theta L_\theta\} \\
&= \frac{1}{2}\operatorname{Tr}\{\Pi_x L_\theta \rho_\theta\} + \frac{1}{2}\operatorname{Tr}\{(\Pi_x \rho_\theta L_\theta)^\dagger\}^* \\
&= \frac{1}{2}\operatorname{Tr}\{\Pi_x L_\theta \rho_\theta\} + \frac{1}{2}\operatorname{Tr}\{L_\theta \rho_\theta \Pi_x\}^* \qquad (44)
\end{aligned}$$

by using the cyclic property of the trace, we can write

$$\partial_\theta p(x|\theta) = \operatorname{Re}(\operatorname{Tr}\{\rho_\theta \Pi_x L_\theta\}) \qquad (45)$$

Then we can use this result to write the Fisher information as

$$F(\theta) = \int \mathrm{d}x\, \frac{\operatorname{Re}(\operatorname{Tr}\{\rho_\theta \Pi_x L_\theta\})^2}{\operatorname{Tr}\{\rho_\theta \Pi_x\}} \qquad (46)$$

For a given quantum measurement, i.e. a POVM $\{\Pi_x\}$, Eq. (41) and Eq. (46) establish the classical bound on precision, which may be achieved by a proper processing.

Quantum Fisher Information and Quantum Cramér-Rao bound

In order to evaluate the ultimate bounds to precision we have now to maximize the Fisher information over the quantum measurements[8]

$$F(\theta) = \int \mathrm{d}x\, \frac{\operatorname{Re}(\operatorname{Tr}\{\rho_\theta \Pi_x L_\theta\})^2}{\operatorname{Tr}\{\rho_\theta \Pi_x\}} \le \frac{1}{\operatorname{Tr}\{\rho_\theta \Pi_x\}} \int \mathrm{d}x \Big| \operatorname{Tr}\{\rho_\theta \Pi_x L_\theta\}\Big|^2 \qquad (47)$$

$$= \frac{1}{\operatorname{Tr}\{\rho_\theta \Pi_x\}} \int \mathrm{d}x \left| \operatorname{Tr}\left\{\Big(\sqrt{\rho_\theta}\sqrt{\Pi_x}\Big)\Big(\sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta}\Big)\right\}\right|^2 \qquad (48)$$

By using the Schwartz inequality:

$$|\operatorname{Tr}(A^\dagger B)|^2 \le \operatorname{Tr}(A^\dagger A)\operatorname{Tr}(B^\dagger B) \qquad (49)$$

we can write

$$\begin{aligned}
\left| \operatorname{Tr}\left\{\Big(\sqrt{\Pi_x}\sqrt{\rho_\theta}\Big)^\dagger \Big(\sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta}\Big)\right\}\right|^2 &\le \operatorname{Tr}\left\{\Big[\sqrt{\Pi_x}\sqrt{\rho_\theta}\Big]^\dagger \Big[\sqrt{\Pi_x}\sqrt{\rho_\theta}\Big]\right\} \times \\
&\times \operatorname{Tr}\left\{[\sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta}]^\dagger [\sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta}]\right\} \\
&= \operatorname{Tr}\left\{\rho_\theta \Pi_x\right\} \operatorname{Tr}\left\{L_\theta \Pi_x L_\theta \rho_\theta\right\} \qquad (50)
\end{aligned}$$

By using Eq. (50) in Eq. (48), we obtain

$$F(\theta) \le \int \mathrm{d}x \operatorname{Tr}\left\{\Pi_x L_\theta \rho_\theta L_\theta\right\} = \operatorname{Tr}\left\{\cancelto{1}{\Big(\int \mathrm{d}x \Pi_x\Big)} L_\theta \rho_\theta L_\theta\right\} \qquad (51)$$

$$F(\theta) \le \operatorname{Tr}\left\{\rho_\theta L_\theta^2\right\} \qquad (52)$$

[8] In Eq. (47) we are using $|z| \ge \operatorname{Re}(z)$.

The above chain of inequalities prove that the Fisher information $F(\theta)$ of any quantum measurement is bounded by the so-called **Quantum Fisher Information** (QFI)[9]

$$F(\theta) \leq H(\theta) \equiv \text{Tr}\left\{\rho_\theta L_\theta^2\right\} = \text{Tr}\left\{(\partial_\theta \rho_\theta) L_\theta\right\} \tag{53}$$

Leading the **quantum Cramér-Rao bound**

$$\boxed{V(\theta) \geq \frac{1}{nH(\theta)}} \tag{54}$$

to the variance of any estimator. The quantum version of the Cramér-Rao theorem provides an ultimate bound.

Optimal POVM

The quantum Fisher Information is an upper bound for the Fisher Information as it embodies the optimization of the Fisher Information over any possible measurement. Optimal quantum measurements for the estimation of θ thus correspond to POVM with Fisher information equal to the quantum Fisher information, i.e. those saturating both inequalities Eq. (47) and Eq. (50). The inequality Eq. (47) is saturated when $\text{Tr}[\rho_\theta \Pi_x L_\theta]$ is a real number. The inequality Eq. (50) is based on the Schwartz inequality

$$|\,\text{Tr}(A^\dagger B)|^2 \leq \text{Tr}(A^\dagger A)\,\text{Tr}(B^\dagger B)$$

Which is saturated when, e.g. $B = cA$ (where c is a constant):

$$|c|^2|\,\text{Tr}(A^\dagger A)|^2 = |c|^2\,\text{Tr}(A^\dagger A)\,\text{Tr}(A^\dagger A)$$

In our case, we shall have

$$\sqrt{\Pi_x}\,\sqrt{\rho_\theta} = c\,\sqrt{\Pi_x} L_\theta\,\sqrt{\rho_\theta} \tag{55}$$

The condition Eq. (55) is satisfied iff $\{\Pi_x\}$ is made by the set of projectors over the states of L_θ[10], which, in turn, represents the optimal POVM to estimate the parameter θ.

Notice, however, that L_θ itself may not represent the optimal observable to be measured. In fact, Eq. (55) **determines the POVM and not the estimator**, i.e. the function of the eigenvalues of L_θ. This corresponds to a classical post-processing of data aimed to saturate the Cramér-Rao inequality and may be pursed by maximum likelihood (see Sec. 1.6.2).

---

[9] In Eq. (53) we are using

$$\frac{L_\theta \rho_\theta + \rho_\theta L_\theta}{2} = \frac{\partial \rho_\theta}{\partial \theta} \implies L_\theta \rho_\theta + \rho_\theta L_\theta = 2\frac{\partial \rho_\theta}{\partial \theta} \implies \left(L_\theta \rho_\theta + \rho_\theta L_\theta\right) L_\theta = (2\partial_\theta \rho_\theta) L_\theta$$

$$\text{Tr}\left\{\left(L_\theta \rho_\theta + \rho_\theta L_\theta\right) L_\theta\right\} = \text{Tr}\left\{(2\partial_\theta \rho_\theta) L_\theta\right\} \implies \text{Tr}\left\{\rho_\theta L_\theta^2\right\} = \text{Tr}\left\{(\partial_\theta \rho_\theta) L_\theta\right\}$$

[10] For instance, if $L_\theta = \sum w_x |q_x\rangle\langle q_x|$, we could have $\{\Pi_x\} = \{|q_x\rangle\langle q_x|\}$. Thus

$$\sqrt{\Pi_x}\,\sqrt{\rho_\theta} = c\,\sqrt{\Pi_x} L_\theta\,\sqrt{\rho_\theta} \implies \sqrt{\Pi_x}\,\sqrt{\rho_\theta} = c\omega_x\,\sqrt{\Pi_x}\,\sqrt{\rho_\theta} \implies \sqrt{\Pi_x}\,\sqrt{\rho_\theta} = \sqrt{\Pi_x}\,\sqrt{\rho_\theta} \text{ with } c = 1/w_x$$

2.2 Expressions for the Symmetric Logarithmic Derivative and for the Quantum Fisher Information

The Eq. (43) is the Lyapunov matrix equation to be solved for L_θ. The general solution may be written as

$$L_\theta = 2\int_0^\infty \mathrm{d}t \exp\Big\{-\rho_\theta t\}(\partial_\theta\rho_\theta)\exp\{-\rho_\theta t\} \tag{56}$$

Proof. Let us cheek if the Eq. (56) is the solution of Eq. (43). Let us start by writing the density matrix in its eigenbasis $\rho_\theta = \sum_n c_n|\psi_n\rangle\langle\psi_n|$. Then we have

$$\begin{aligned}\rho_\theta L_\theta &= 2\Big(\sum_i c_i|\psi_i\rangle\langle\psi_i|\Big)\int_0^\infty \mathrm{d}t\Big(\sum_n e^{-c_n t}|\psi_n\rangle\langle\psi_n|\Big)(\partial_\theta\rho_\theta)\times\\ &\times\Big(\sum_m e^{-c_m t}|\psi_m\rangle\langle\psi_m|\Big)\\ &= \sum_{i,n,m} 2c_i\langle\psi_i|\psi_n\rangle\int_0^\infty \mathrm{d}t e^{-(c_n+c_m)t}\langle\psi_n|(\partial_\theta\rho_\theta)|\psi_m\rangle|\psi_i\rangle\langle\psi_m|\\ &= \sum_{n,m}\frac{2c_n}{c_n+c_m}\langle\psi_n|(\partial_\theta\rho_\theta)|\psi_m\rangle|\psi_n\rangle\langle\psi_m| \end{aligned}\tag{57}$$

we also have

$$\begin{aligned}(L_\theta\rho_\theta)^\dagger = L_\theta\rho_\theta &= \sum_{n,m}\frac{2c_n}{c_n+c_m}\langle\psi_m|(\partial_\theta\rho_\theta)|\psi_n\rangle|\psi_m\rangle\langle\psi_n|\\ &= \sum_{n,m}\frac{2c_m}{c_n+c_m}\langle\psi_n|(\partial_\theta\rho_\theta)|\psi_m\rangle|\psi_n\rangle\langle\psi_m| \end{aligned}\tag{58}$$

Combining Eq. (57) and Eq. (58), we obtain

$$\begin{aligned}\frac{L_\theta\rho_\theta+\rho_\theta L_\theta}{2} &= \sum_{n,m}\Big(\frac{c_n+c_m}{c_n+c_m}\Big)\langle\psi_n|(\partial_\theta\rho_\theta)|\psi_m\rangle|\psi_n\rangle\langle\psi_m|\\ &= \sum_{n,m}\langle\psi_n|(\partial_\theta\rho_\theta)|\psi_m\rangle|\psi_n\rangle\langle\psi_m|\\ &= \frac{\partial}{\partial\theta}\sum_{n,m}\langle\psi_n|\rho_\theta|\psi_m\rangle|\psi_n\rangle\langle\psi_m|\\ &= \frac{\partial}{\partial\theta}\sum_{n,m}|\psi_n\rangle\langle\psi_n|\rho_\theta|\psi_m\rangle\langle\psi_m|\\ &= \partial_\theta\rho_\theta \end{aligned}\tag{59}$$

□

Upon writing ρ_θ in its eigenbasis $\rho_\theta = \sum_n c_n|\psi_n\rangle\langle\psi_n|$, leads to

$$\begin{aligned}L_\theta &= 2\sum_{n,m}\int_0^\infty \mathrm{d}t\, e^{-(c_n+c_m)t}|\psi_n\rangle\langle\psi_n|(\partial_\theta\rho_\theta)|\psi_m\rangle\langle\psi_m|\\ &= 2\sum_{n,m}\frac{\langle\psi_n|\partial_\theta\rho_\theta|\psi_m\rangle}{c_n+c_m}|\psi_n\rangle\langle\psi_m| \end{aligned}\tag{60}$$

By using the Eq. (60) we can write the quantum Fisher information as,

$$
\begin{aligned}
H(\theta) &= \mathrm{Tr}\left\{(\partial_\theta\rho_\theta)L_\theta\right\} \\
&= \mathrm{Tr}\left\{(\partial_\theta\rho_\theta)\Big(2\sum_{n,m}\frac{\langle\psi_n|\partial_\theta\rho_\theta|\psi_m\rangle}{c_n+c_m}|\psi_n\rangle\langle\psi_m|\Big)\right\} \\
&= 2\sum_{n,m}\frac{\langle\psi_n|\partial_\theta\rho_\theta|\psi_m\rangle}{c_n+c_m}\,\mathrm{Tr}\left\{\partial_\theta\rho_\theta|\psi_n\rangle\langle\psi_m|\right\} \\
&= 2\sum_{n,m}\frac{\langle\psi_n|\partial_\theta\rho_\theta|\psi_m\rangle}{c_n+c_m}\langle\psi_m|\partial_\theta\rho_\theta|\psi_n\rangle \\
H(\theta) &= 2\sum_{n,m}\frac{|\langle\psi_n|\partial_\theta\rho_\theta|\psi_m\rangle|^2}{c_n+c_m} \qquad (61)
\end{aligned}
$$

2.3 Classical and Quantum contributions to the Quantum Fisher Information

Notice that the SLD is defined only on the support of ρ_θ [see Eq. (60)] and both the eigenvalues c_n and the eigenvectors $|\psi_n\rangle$ may depend on the parameter. In order to separate the two contributions to the quantum Fisher information we explicitly evaluate $\partial_\theta\rho_\theta$

$$
\begin{aligned}
\partial_\theta\rho_\theta &= \partial_\theta\Big(\sum_n c_n|\psi_n\rangle\langle\psi_n|\Big) \\
&= \sum_n\Big[(\partial_\theta c_n)|\psi_n\rangle\langle\psi_n| + c_n|\partial_\theta\psi_n\rangle\langle\psi_n| + c_n|\psi_n\rangle\langle\partial_\theta\psi_n|\Big] \qquad (62)
\end{aligned}
$$

Here we are using the notation

$$
|\partial_\theta\psi_n\rangle = \partial_\theta|\psi_n\rangle = \sum_k \partial_\theta\psi_{nk}|k\rangle \qquad (63)
$$

where ψ_{nk} are obtained expanding $|\psi_n\rangle$ in arbitrary basis $\{|k\rangle\}$ independent on θ. Since $\langle\psi_n|\psi_m\rangle = \delta_{nm}$ we have

$$
\partial_\theta\Big[\langle\psi_n|\psi_m\rangle\Big] = \langle\partial_\theta\psi_n|\psi_m\rangle + \langle\psi_n|\partial_\theta\psi_m\rangle = 0 \Longrightarrow \boxed{\langle\partial_\theta\psi_n|\psi_m\rangle = -\langle\psi_n|\partial_\theta\psi_m\rangle} \qquad (64)
$$

Using Eqs. (60) and (62) we have

$$
\begin{aligned}
L_\theta &= \sum_{n,m,i}\frac{2}{c_n+c_m}\langle\psi_n|\Big[(\partial_\theta c_i)|\psi_i\rangle\langle\psi_i| + c_i|\partial_\theta\psi_i\rangle\langle\psi_i| + c_i|\psi_i\rangle\langle\partial_\theta\psi_i|\Big]|\psi_m\rangle|\psi_n\rangle\langle\psi_m| \\
&= \sum_n\frac{\partial_\theta c_n}{c_n}|\psi_n\rangle\langle\psi_n| + \sum_{n,m,i}\frac{2c_i}{c_n+c_m}\Big[\langle\psi_n|\partial_\theta\psi_i\rangle\langle\psi_i|\psi_m\rangle + \langle\psi_n|\psi_i\rangle\langle\partial_\theta\psi_i|\psi_m\rangle\Big]|\psi_n\rangle\langle\psi_m| \\
&= \sum_n\frac{\partial_\theta c_n}{c_n}|\psi_n\rangle\langle\psi_n| + \sum_{n,m}\frac{2}{c_n+c_m}\Big[c_m\langle\psi_n|\partial_\theta\psi_m\rangle + c_n\langle\partial_\theta\psi_n|\psi_m\rangle\Big]|\psi_n\rangle\langle\psi_m| \qquad (65)
\end{aligned}
$$

Now, we can use the result Eq. (64)

$$L_\theta = \sum_n \frac{\partial_\theta c_n}{c_n}|\psi_n\rangle\langle\psi_n| + \sum_{n,m}\frac{2}{c_n+c_m}\Big[c_m\langle\psi_n|\partial_\theta\psi_m\rangle - c_n\langle\psi_n|\partial_\theta\psi_m\rangle\Big]|\psi_n\rangle\langle\psi_m|$$
$$= \sum_n \frac{\partial_\theta c_n}{c_n}|\psi_n\rangle\langle\psi_n| + 2\sum_{n,m}\frac{c_m-c_n}{c_n+c_m}\langle\psi_n|\partial_\theta\psi_m\rangle|\psi_n\rangle\langle\psi_m| \tag{66}$$

Squaring the above expression, we find

$$L_\theta^2 = \Big(\sum_n \frac{\partial_\theta c_n}{c_n}|\psi_n\rangle\langle\psi_n| + 2\sum_{n,m}\frac{c_m-c_n}{c_n+c_m}\langle\psi_n|\partial_\theta\psi_m\rangle|\psi_n\rangle\langle\psi_m|\Big)\times$$
$$\Big(\sum_{n'} \frac{\partial_\theta c_{n'}}{c_{n'}}|\psi_{n'}\rangle\langle\psi_{n'}| + 2\sum_{n',m'}\frac{c_{m'}-c_{n'}}{c_{n'}+c_{m'}}\langle\psi_{n'}|\partial_\theta\psi_{m'}\rangle|\psi_{n'}\rangle\langle\psi_{m'}|\Big)$$
$$= \sum_n\Big(\frac{\partial_\theta c_n}{c_n}\Big)^2|\psi_n\rangle\langle\psi_n| + 2\sum_{n,m'}\frac{\partial_\theta c_n}{c_n}\Big(\frac{c_{m'}-c_n}{c_n+c_{m'}}\Big)\langle\psi_n|\partial_\theta\psi_{m'}\rangle|\psi_n\rangle\langle\psi_{m'}|$$
$$+2\sum_{n,m}\frac{\partial_\theta c_m}{c_m}\Big(\frac{c_m-c_n}{c_n+c_m}\Big)\langle\psi_n|\partial_\theta\psi_m\rangle|\psi_n\rangle\langle\psi_m|$$
$$+4\sum_{n,m,m'}\Big(\frac{c_m-c_n}{c_n+c_m}\Big)\Big(\frac{c_{m'}-c_m}{c_m+c_{m'}}\Big)\langle\psi_n|\partial_\theta\psi_m\rangle\langle\psi_m|\partial_\theta\psi_{m'}\rangle|\psi_n\rangle\langle\psi_{m'}| \tag{67}$$

We can use the above result to calculate the quantum Fisher information

$$H(\theta) = \mathrm{Tr}\{\rho_\theta L_\theta^2\} = \mathrm{Tr}\left\{\Big(\sum_n c_n|\psi_n\rangle\langle\psi_n|\Big)L_\theta^2\right\} = \sum_n c_n\langle\psi_n|L_\theta^2|\psi_n\rangle \tag{68}$$

Using the expression for L_θ^2 we have

$$H(\theta) = \sum_n\frac{(\partial_\theta c_n)^2}{c_n} + 4\sum_{n,m}c_n\Big(\frac{c_m-c_n}{c_n+c_m}\Big)\Big(\frac{c_n-c_m}{c_m+c_n}\Big)\langle\psi_n|\partial_\theta\psi_m\rangle\langle\psi_m|\partial_\theta\psi_n\rangle \tag{69}$$

Thus, we can write[11]

$$H(\theta) = \sum_n\frac{(\partial_\theta c_n)^2}{c_n} + 2\sum_{n,m}\tilde{\sigma}_{nm}|\langle\psi_n|\partial_\theta\psi_m\rangle|^2 \tag{70}$$

with $\tilde{\sigma}_{nm} = 2c_n[(c_m-c_n)/(c_n+c_m)]^2$. Note that we can also write a more general expression

$$H(\theta) = \sum_n\frac{(\partial_\theta c_n)^2}{c_n} + 2\sum_{n\neq m}(\tilde{\sigma}_{nm}+\Lambda_{nm})|\langle\psi_n|\partial_\theta\psi_m\rangle|^2 \tag{71}$$

where Λ_{nm} is any antisymmetric term[12], i.e, $\Lambda_{nm} = -\Lambda_{mn}$. If we define $\sigma_{nm} = \tilde{\sigma}_{nm} + \Lambda_{nm}$, we have

$$\boxed{H(\theta) = \underbrace{\sum_n\frac{(\partial_\theta c_n)^2}{c_n}}_{\substack{\text{classical}\\\text{Fisher information}}} + \underbrace{2\sum_{n\neq m}\sigma_{nm}|\langle\psi_n|\partial_\theta\psi_m\rangle|^2}_{\text{truly quantum contribution}}} \tag{72}$$

[11] In Eq. (70) we are using the result Eq. (64), i.e. $\langle\psi_m|\partial_\theta\psi_n\rangle = -\langle\partial_\theta\psi_m|\psi_n\rangle$

[12] It is easy see why that is true. Again, just remember that $\langle\delta_\theta\psi_n|\psi_m\rangle = -\langle\psi_n|\delta_\theta\psi_m\rangle$

Some examples of possible values for σ_{nm} are

$$\sigma_{nm} = \tilde{\sigma}_{nm} + \cancelto{0}{\Delta_{nm}} = 2c_n\left(\frac{c_m - c_n}{c_n + c_m}\right)^2 \tag{73}$$

$$\sigma_{nm} = \tilde{\sigma}_{nm} + 2\frac{(c_m - c_n)^3}{(c_m + c_n)^2} = 2c_m\left(\frac{c_m - c_n}{c_n + c_m}\right)^2 \tag{74}$$

$$\sigma_{nm} = \tilde{\sigma}_{nm} + \frac{2(c_m - c_n)(c_m^2 + c_m^2)}{(c_m + c_n)^2} = 2c_m\left(\frac{c_m - c_n}{c_n + c_m}\right) \tag{75}$$

$$\sigma_{nm} = \tilde{\sigma}_{nm} + \frac{(c_m - c_n)^3}{(c_m + c_n)^2} = \frac{(c_m - c_n)^2}{c_n + c_m} \tag{76}$$

The first term in Eq. (72) represents the classical Fisher information of the distribution c_n $\Big[$remember that $\rho_\theta = \sum_n c_n|\psi_n\rangle\langle\psi_n|$ $\Big]$ whereas the second term contains the truly quantum contribution.

2.4 Unitary families and the pure state model

Let us consider the case where the parameter of interest is the amplitude of a unitary perturbation imposed to a given initial state ρ_0. The family of quantum states we are dealing with may be expressed as

$$\rho_\theta = U_\theta \rho_0 U_\theta^\dagger \tag{77}$$

where $U_\theta = \exp\{-i\theta G\}$ is a unitary operator and G is the corresponding Hermitian generator. Upon expanding the unperturbed state in its eigenbasis

$$\rho_0 = \sum_k c_k|k\rangle\langle k| \quad \text{we have} \quad \rho_\theta = \sum_k c_k|\psi_k\rangle\langle\psi_k| \quad \text{where} \quad |\psi_k\rangle = U_\theta|k\rangle \tag{78}$$

as a consequence we have

$$\begin{aligned}\partial_\theta\rho_\theta &= \sum_k c_k\Big[(\partial_\theta|\psi_k\rangle)\langle\psi_k| + |\psi_k\rangle(\partial_\theta\langle\psi_k|)\Big] = -i\sum_k c_k\Big[G|\psi_k\rangle\langle\psi_k| - |\psi_k\rangle\langle\psi_k|G\Big] \\ &= -i\Big[G,\rho_\theta\Big] = -iU_\theta\Big[G,\rho_0\Big]U_\theta^\dagger \end{aligned} \tag{79}$$

Let us calculate the SLD,

$$\begin{aligned} L_\theta &= 2\sum_{n,m}\frac{\langle\psi_n|\partial_\theta\rho_\theta|\psi_m\rangle}{c_n + c_m}|\psi_n\rangle\langle\psi_m| \\ &= -2i\sum_{n,m}\frac{\langle n|U_\theta^\dagger U_\theta\Big[G,\rho_0\Big]U_\theta^\dagger U_\theta|m\rangle}{c_n + c_m}U_\theta|n\rangle\langle m|U_\theta^\dagger \\ &= U_\theta\Big(-2i\sum_{n,m}\frac{\langle n|\Big[G,\rho_0\Big]|m\rangle}{c_n + c_m}|n\rangle\langle m|\Big)U_\theta^\dagger \\ &= U_\theta\Big(-2i\sum_{n,m}\langle n|G|m\rangle\Big(\frac{c_m - c_n}{c_n + c_m}\Big)|n\rangle\langle m|\Big)U_\theta^\dagger \end{aligned} \tag{80}$$

If we define

$$L_0 = -2i \sum_{n,m} \langle n|G|m\rangle \Big(\frac{c_m - c_n}{c_n + c_m}\Big) |n\rangle\langle m| \quad \text{we can write} \quad L_\theta = U_\theta L_0 U_\theta^\dagger \tag{81}$$

The corresponding quantum Fisher information is independent on the value of the parameter and may be written in compact form as

$$H(\theta) = \text{Tr}\{\rho_\theta L_\theta^2\} = \text{Tr}\{(U_\theta \rho_0 U_\theta^\dagger)(U_\theta L_0^2 U_\theta^\dagger)\} = \text{Tr}\{\rho_0 L_0^2\} \tag{82}$$

Thus we have to calculate L_0^2

$$\begin{aligned} L_0^2 &= \Big[-2i \sum_{n,m} \langle n|G|m\rangle \Big(\frac{c_m - c_n}{c_n + c_m}\Big) |n\rangle\langle m| \Big]\Big[-2i \sum_{n',m'} \langle n'|G|m'\rangle \Big(\frac{c_{m'} - c_{n'}}{c_{n'} + c_{m'}}\Big) |n'\rangle\langle m'| \Big] \\ &= -4 \sum_{n,m,m'} \langle n|G|m\rangle\langle m|G|m'\rangle \Big(\frac{c_m - c_n}{c_n + c_m}\Big)\Big(\frac{c_{m'} - c_m}{c_m + c_{m'}}\Big) |n\rangle\langle m'| \end{aligned} \tag{83}$$

and $\rho_0 L_0^2$

$$\begin{aligned} \rho_0 L_0^2 &= -4 \Big(\sum_k c_k |k\rangle\langle k|\Big) \sum_{n,m,m'} \langle n|G|m\rangle\langle m|G|m'\rangle \Big(\frac{c_m - c_n}{c_n + c_m}\Big)\Big(\frac{c_{m'} - c_m}{c_m + c_{m'}}\Big) |n\rangle\langle m'| \\ &= -4 \sum_{n,m,m'} c_n \langle n|G|m\rangle\langle m|G|m'\rangle \Big(\frac{c_m - c_n}{c_n + c_m}\Big)\Big(\frac{c_{m'} - c_m}{c_m + c_{m'}}\Big) |n\rangle\langle m'| \end{aligned} \tag{84}$$

Finally we have

$$H(\theta) = \text{Tr}\{\rho_0 L_0^2\} = 2 \sum_{n,m} \sigma_{nm} |\langle n|G|m\rangle|^2 \tag{85}$$

where $\sigma_{nm} = 2c_n((c_m - c_n)/(c_n + c_m))^2$. Here again we can write $\sigma_{nm} = [\sigma_{nm} +$ any antisymmetric]. Possible values for σ_{nm} are, for example, Eqs. (73)-(76).

Pure state model

For a generic family of pure states we have $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$. Since $\rho_\theta^2 = \rho_\theta$ we have

$$\partial_\theta \rho_\theta = \partial_\theta(\rho_\theta^2) = \rho_\theta(\partial_\theta \rho_\theta) + (\partial_\theta \rho_\theta)\rho_\theta \tag{86}$$

If we compare Eq. (43) and Eq. (86), we find

$$L_\theta = 2\partial_\theta \rho_\theta = 2\partial_\theta(|\psi_\theta\rangle\langle\psi_\theta|) = 2\Big[|\partial_\theta\psi_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\Big] \tag{87}$$

Finally we can calculate the Fisher Information[13]

$$\begin{aligned}
H(\theta) &= \mathrm{Tr}\left\{(\partial_\theta\rho_\theta)L_\theta\right\} = \frac{1}{2}\,\mathrm{Tr}\left\{L_\theta^2\right\} \\
&= 2\,\mathrm{Tr}\left\{\Big[|\partial_\theta\psi_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\Big]\Big[|\partial_\theta\psi_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\Big]\right\} \\
&= 2\,\mathrm{Tr}\Big\{|\partial_\theta\psi_\theta\rangle\langle\psi_\theta|\partial_\theta\psi_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\partial_\theta\psi_\theta\rangle\langle\psi_\theta| + \\
&\qquad\qquad + |\partial_\theta\psi_\theta\rangle\langle\psi_\theta|\psi_\theta\rangle\langle\partial_\theta\psi_\theta| + |\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\Big\} \\
&= 2\Big[\langle\psi_\theta|\partial_\theta\psi_\theta\rangle\langle\psi_\theta|\partial_\theta\psi_\theta\rangle + \langle\partial_\theta\psi_\theta|\partial_\theta\psi_\theta\rangle\langle\psi_\theta|\psi_\theta\rangle + \\
&\qquad\qquad + \langle\psi_\theta|\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\partial_\theta\psi_\theta\rangle + \langle\partial_\theta\psi_\theta|\psi_\theta\rangle\langle\partial_\theta\psi_\theta|\psi_\theta\rangle\Big] \\
&= 4\Big[\langle\partial_\theta\psi_\theta|\partial_\theta\psi_\theta\rangle + \langle\partial_\theta\psi_\theta|\psi_\theta\rangle^2\Big] \qquad (88)
\end{aligned}$$

For a unitary family of pure states $|\psi_\theta\rangle = U_\theta|\psi_0\rangle$ we have the following results:

$$|\partial_\theta\psi_\theta\rangle = \partial_\theta\big(U_\theta|\psi_0\rangle\big) = \big(\partial_\theta U_\theta\big)|\psi_0\rangle = -iGU_\theta|\psi_0\rangle = -iG|\psi_\theta\rangle \qquad (89)$$

$$\langle\partial_\theta\psi_\theta|\partial_\theta\psi_\theta\rangle = \langle\psi_\theta|G^2|\psi_\theta\rangle = \langle\psi_0|U_\theta^\dagger G^2 U_\theta|\psi_0\rangle = \langle\psi_0|G^2|\psi_0\rangle \qquad (90)$$

$$\langle\partial_\theta\psi_\theta|\psi_\theta\rangle = i\langle\psi_\theta|G|\psi_\theta\rangle = i\langle\psi_0|U_\theta^\dagger G U_\theta|\psi_0\rangle = i\langle\psi_0|G|\psi_0\rangle \qquad (91)$$

The quantum Fisher information thus reduces to the simple form

$$\begin{aligned}
H(\theta) &= 4\Big[\langle\psi_0|G^2|\psi_0\rangle + (i\langle\psi_0|G|\psi_0\rangle)^2\Big] \\
&= 4\Big[\langle\psi_0|G^2|\psi_0\rangle - \langle\psi_0|G|\psi_0\rangle^2\Big] \\
&= 4\langle\psi_0|(\Delta G)^2|\psi_0\rangle \\
&= 4\langle(\Delta G)^2\rangle \qquad (92)
\end{aligned}$$

where we are using the definition $\langle\psi_0|(\Delta G)^2|\psi_0\rangle = \langle\psi_0|G^2|\psi_0\rangle - \langle\psi_0|G|\psi_0\rangle^2$. The quantum Fisher information is independent on θ and proportional to the fluctuations or the generator on the unperturbed state. Using the quantum Crammér-Rao bound, we have

$$V(\theta) \geq \frac{1}{nH(\theta)} \quad \Longrightarrow \quad V(\theta) \geq \frac{1}{4n\langle(\Delta G)^2\rangle} \qquad (93)$$

Mixed state

We already calculate the QFI for a mixed state in Eq. (85). Here we intend to recast that result in a special form. Let us start by writing

$$H(\theta) = 2\sum_{n,m}\sigma_{nm}|\langle n|G|m\rangle|^2 = \sum_{n,m}4c_n\left(\frac{c_m - c_n}{c_n + c_m}\right)\langle n|G|m\rangle\langle m|G|n\rangle \qquad (94)$$

We would like to rewrite this result as a function of the variance $\langle(\Delta G)^2\rangle$,

$$\langle(\Delta G)^2\rangle = \mathrm{Tr}\left\{(\Delta G)^2\rho_0\right\} = \langle G^2\rangle - \langle G\rangle^2 = \mathrm{Tr}\left\{G^2\rho_0\right\} - \mathrm{Tr}\left\{G\rho_0\right\}^2 \qquad (95)$$

[13] Here we will use again the result $\langle\partial_\theta\psi_\theta|\psi_\theta\rangle = -\langle\psi_\theta|\partial_\theta\psi_\theta\rangle$.

First, note that we can write

$$\mathrm{Tr}\left\{G^2\rho_0\right\} = \mathrm{Tr}\left\{G^2\Big(\sum_n c_n|n\rangle\langle n|\Big)\right\} = \sum_n c_n\langle n|G^2|n\rangle$$
$$= \sum_{n,m} c_n\langle n|G|m\rangle\langle m|G|n\rangle \tag{96}$$

Using this result we can write

$$H(\theta) = 4\langle(\Delta G)^2\rangle + 4\langle G\rangle^2 + \sum_{n,m} 4\left[c_n\Big(\frac{c_m - c_n}{c_n + c_m}\Big) - c_n\right]\langle n|G|m\rangle\langle m|G|n\rangle$$
$$= 4\langle(\Delta G)^2\rangle + 4\left[\langle G\rangle^2 + \sum_{n,m} 2c_n\Big(\frac{c_m}{c_n + c_m}\Big)\langle n|G|m\rangle\langle m|G|n\rangle\right]$$
$$= 4\langle(\Delta G)^2\rangle + 4\sum_n c_n\langle n|\left[\langle G\rangle^2 + \sum_{n,m} 2\Big(\frac{c_m}{c_n + c_m}\Big)G|m\rangle\langle m|G\right]|n\rangle \tag{97}$$

Finally, if we define

$$K^{(n)} = \sum_m \frac{c_m}{c_n + c_m}|m\rangle\langle m| \tag{98}$$

we can write

$$H(\theta) = 4\,\mathrm{Tr}\left\{(\Delta G)^2\rho_0\right\} + 4\sum_n c_n\langle n|\left[\langle G\rangle^2 + \sum_n 2GK^{(n)}G\right]|n\rangle \tag{99}$$

Now the quantum Crammér-rao bound, $V(\theta) \geq 1/nH(\theta)$, can be written as

$$V(\theta) \geq \frac{1}{4n}\Bigg[\langle(\Delta G)^2\rangle + \underbrace{\sum_n c_n\langle n|\left[\langle G\rangle^2 + \sum_n 2GK^{(n)}G\right]|n\rangle}_{\substack{\text{classical contribution}\\ \text{due to the mixing}}}\Bigg]^{-1} \tag{100}$$

The second term in Eq. (100) thus represents the classical contribution to uncertainty due to the mixing thus represent the classical contribution to uncertainty due to the mixing of the initial signal.

References

[1] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole, 8th ed., January 2011. ISBN-13: 978-0-538-73352-6.

[2] J. Aldrich, "R. A. Fisher and the making of maximum likelihood 1912-1922," *Statistical Science*, vol. 12, no. 3, pp. 162–176, 1997.

[3] M. G. A. PARIS, "Quantum estimation for quantum technology," *International Journal of Quantum Information*, vol. 07, no. supp01, pp. 125–137, 2009.