Lecture 11: October 10, 2024

Scribe: Louis Marquis *Quantum Sensing and Fisher Information*

11.1 Quantum Sensing

We have previously applied Holevo Information and relative entropy to classical information theory problems like hypothesis testing, channel encoding, and state tomography. We will now apply it to quantum sensing, which is a variant of state tomography.

Suppose that we have a magnetic field with unknown magnitude B and an electron (comprising a qubit), which results in the following Hamiltonian.

$$H = \frac{B}{2} Z \tag{11.1}$$

If there are N such particles, the total Hamiltonian is the sum of the individuals.

$$H = \sum_{i=1}^{N} \frac{B}{2} Z_i \tag{11.2}$$

The goal is to estimate B as precisely as possible. We wish to determine the a informationally-theoretic limit on such a precision.
A first attempt on the one-qubit scenario may involve evolving the state $|+\rangle$ with the Hamiltonian H, then measuring in the $\{|+\rangle, |-\rangle\}$ basis after time t. Denote $\phi = Bt$.

$$e^{-iHt} |+\rangle = \frac{1}{\sqrt{2}} (e^{-\frac{i\phi}{2}} |0\rangle + e^{\frac{i\phi}{2}} |1\rangle) \tag{11.3}$$

$$| \langle +| e^{-iHt} |+\rangle |^2 = | \langle +| \frac{1}{\sqrt{2}} (e^{-\frac{i\phi}{2}} |0\rangle + e^{\frac{i\phi}{2}} |1\rangle)|^2 = \cos^2 \frac{\phi}{2} \tag{11.4}$$

Let random variable X represent the sign of the measured state. That is, $X = 1$ if $|+\rangle$ is measured, and $X = -1$ if $|-\rangle$ is measured. Clearly, the expected value is $E(X) = \cos^2 \frac{\phi}{2} - \sin^2 \frac{\phi}{2} = \cos \phi$. If the experiment is repeated many (N) times, the average X should approach the expected value $\bar{X}$, allowing an estimate ϕ (and in turn B).

$$\bar{x} = \frac{\sum_{i=1}^{N} X_i}{N} \tag{11.5}$$

$$\phi = \arccos \bar{x} \tag{11.6}$$

$$B = \frac{\phi}{t} \tag{11.7}$$

We want to estimate the uncertainty of B. We can calculate the uncertainty Δx, which we define as the standard deviation of x. Then the uncertainty can be propogated over to get ΔB.

$$\Delta x = \sigma(X) = \sqrt{E(X^2) - E(X)^2} = \sqrt{1 - \cos^2 \phi} = |\sin(\phi)| \tag{11.8}$$

$$\Delta\bar{x} = \sigma(\frac{\sum_{i=1}^{N} X_i}{N}) = \frac{\sigma(X)}{\sqrt{N}} = \frac{|\sin(\phi)|}{\sqrt{N}} \tag{11.9}$$

$$\Delta\phi = \frac{\Delta\bar{x}}{|\frac{d\bar{x}}{d\phi}|} = \frac{\frac{|\sin(\phi)|}{\sqrt{N}}}{|\sin\phi|} = \frac{1}{\sqrt{N}} \tag{11.10}$$

$$\Delta B = \frac{\Delta\phi}{t} = \frac{1}{\sqrt{N}t} \tag{11.11}$$

Particularly remarkable about this result is that neither $\Delta\phi$ nor ΔB depend on anything besides B and t, including anything relating to phase. This means that the experiment is equally precise regardless of the phase of the starting state.
$\frac{1}{\sqrt{N}}$ is referred to as the standard quantum limit (SQL), or the shot-noise limit.
This first attempt consisted of independent measurements of N different qubits. A better precision can be achieved by entangling the qubits first into the cat state. In practice, the cat state is more vulnerable to noise but we can ignore this aspect for now. We evolve this state with the N qubit Hamiltonian.

$$e^{-iHt}\frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}) = \frac{1}{\sqrt{2}}(e^{-\frac{i\phi N}{2}}|0\rangle^{\otimes N} + e^{\frac{i\phi N}{2}}|1\rangle^{\otimes N}) \tag{11.12}$$

We notice that the phase shift increases by a factor of N (relative to the first attempt), but this time we only run the multi-qubit experiment once (instead of the single-qubit one N times). This allows a quick calculation of the new precision.

$$\Delta B = \frac{1}{Nt} \tag{11.13}$$

This is referred to as the Heisenberg limit, as it parallels the Heisenberg Uncertainty Principle.

11.2 Hamiltonian Learning

The problem of quantum sensing is a specific case of the more general problem of Hamiltonian learning, in which the Hamiltonian. In theory, the maximally general Hamiltonian can have 4^N unknown parameters, but such a problem isn't very useful nor interesting. In Hamiltonian learning, we know the general structure of the Hamiltonian as the linear combination of a set of terms $\{h_i\}$ and seek the specific parameters $\{b_i \in \mathbb{R}\}$ of this combination.

$$H = \sum_i \beta_i h_i \tag{11.14}$$

In this problem, one must prepare a state, evolve it with the Hamiltonian, and measure it to gain information on $\{h_i\}$.
Alternatively, one might be given the Gibbs state for a Hamiltonian $\frac{e^{-\frac{H}{T}}}{\mathrm{tr}\left(e^{-\frac{H}{T}}\right)}$ and must determine $\{h_i\}$ by measuring the state. This problem is also referred to as Hamiltonian learning.
However, Hamiltonian learning is a very complicated topic since there are so many strategies that can be considered. So for the rest of this lecture, we will focus on a simpler problem called parameter estimation from states. This problem is simple enough to have a full solution, and this solution reveals insights on Hamiltonian learning.

11.3 Parameter Estimation from States

In this problem, there is an unknown parameter θ that determines the distribution $p_\theta(x)$ of observation x. The goal is to output an optimal estimate $\hat{\theta}$ after observing x. It is assumed that $p_\theta(x)$ is continuous and differentiable over θ.
A simpler version of the problem involves distinguishing $p_\theta(x)$ from $p_0(x)$ for θ close to 0. In this case, we can simply use a likelihood ratio test for hypothesis testing. Define W_n as the logarithm of such a ratio.

$$W_n(x^n) = \log \frac{\prod_{i=1}^n p_\theta(x_i)}{\prod_{i=1}^n p_0(x_i)} \tag{11.15}$$

We have seen before a bound on the expectation of W_n for the $x^n \leftarrow p_0^n$ case.

$$E_{x^n \leftarrow p_0^n}(W_n) \leq 0 \tag{11.16}$$

We can also calculate the $x^n \leftarrow p_\theta^n$ case.

$$E_{x^n \leftarrow p_\theta^n}(W_n) = \sum_{x^n} p_\theta^n(x^n) \log \frac{\prod_{i=1}^n p_\theta(x_i)}{\prod_{i=1}^n p_0(x_i)} \tag{11.17}$$

$$= \sum_{i=1}^n \sum_{x^n} p_\theta^n(x^n) \log \frac{p_\theta(x_i)}{p_0(x_i)} \tag{11.18}$$

$$= \sum_{i=1}^n \sum_{x_i} p_\theta(x_i) \log \frac{p_\theta(x_i)}{p_0(x_i)} \tag{11.19}$$

$$= nD(p_\theta || p_0) \tag{11.20}$$

We can estimate the relative entropy $D(p_\theta||p_0)$ for small θ by expanding it as a power series. At $\theta = 0$, we know that it is zero and symmetric, so the constant and linear terms must be zero. Therefore, the first potentially nonzero term is the quadratic term, which we denote F.

$$E(W_n) = nD(p_\theta||p_0) = n(0 * 1 + 0 * \theta + F * \frac{\theta^2}{2} + O(\theta^3)) = \frac{nF\theta^2}{2} + O(\theta^3) \tag{11.21}$$

We can restate this a direct formula for F, which we call the Fisher information.

$$F = \partial_\theta^2 D(p_\theta||p_0)|_{\theta=0} \tag{11.22}$$

The Fisher information has several equivalent forms, which are useful but will not be derived in this lecture.

$$F = \sum_x p_\theta(x)(\partial_\theta \log p_\theta(x))^2|_{\theta=0} \tag{11.23}$$

$$= \sum_x \frac{(\partial_\theta p_\theta(x))^2}{p_\theta}|_{\theta=0} \tag{11.24}$$

We now show that the Fisher information indicates whether p_0^n and p_θ^n can be reliably distinguished. We define this condition as the expectation of W_n under p_θ^n being greater than its standard deviation under either p_θ^n or p_0^n. Such a condition can be found by calculating the variance with a factor of

$\frac{1}{n}$ for convenience.

$$\frac{1}{n}\sigma^2_{x^n \leftarrow p_\theta^n}(W_n) = \sum_x p_\theta(x)(\log\frac{p_\theta(x)}{p_0(x)})^2 - D(p_\theta, p_0)^2 \tag{11.25}$$

$$= \sum_x p_\theta(x)(\theta\partial_\theta \log p_\theta + O(\theta^2))^2 - O(\theta^4) \tag{11.26}$$

$$= \theta^2 \sum_x p_\theta(x)(\partial_\theta \log p_\theta)^2 + O(\theta^3) \tag{11.27}$$

$$= F\theta^2 + O(\theta^3) \tag{11.28}$$

To reliably distinguish the two distributions, the expectation must greater than the standard deviation. We ignore $O(\theta^3)$ elements.

$$\frac{nF\theta^2}{2} = E(W_n) \geq \sqrt{\sigma^2(W_n)} = \sqrt{nF\theta^2} \implies \theta \geq \frac{1}{\sqrt{nF}} \tag{11.29}$$

This is the minimum θ at which one can reliably distinguish p_θ, p_0.

11.4 Cramer-Rao Bound

We have found the minimum θ whose distribution can be reliably distinguished from that of 0. This quantity also happends to also be the Cramer-Rao Bound.

Theorem 11.4.1. *Define an estimator $\hat{\theta}(x^n)$ as locally unbiased if the expectation of the estimate $\hat{\theta}$ is approximately θ when θ is close to θ_0. θ_0 will almost always be set to 0 in practice.*

$$E_{x^n \leftarrow p_\theta^n}(\hat{\theta}) = \theta + O((\theta - \theta_0)^2) \tag{11.30}$$

If $\hat{\theta}(x^n)$ is locally unbiased, then $\sigma(\hat{\theta}) \geq \frac{1}{\sqrt{nF}}$.

Proof. We wish to find a lower bound for the variance of $\hat{\theta}$. We can begin by noticing that $E(\hat{\theta}) = \theta + O(\theta^2)$ and since θ is small, this term can be ignored in the variance of $\hat{\theta}$.

$$\sigma^2(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2 \approx E(\hat{\theta}^2) = \sum_{x^n} p_\theta^n(x^n)\hat{\theta}(x^n)^2 \tag{11.31}$$

We can then take a derivative of the locally unbiased condition.

$$1 = \partial_\theta E(\hat{\theta})|_{\theta=0} \tag{11.32}$$

$$= \partial_\theta \sum_{x^n} p_\theta^n(x^n)\hat{\theta}(x^n)|_{\theta=0} \tag{11.33}$$

$$= \sum_{x^n} \partial_\theta p_\theta^n(x^n)|_{\theta=0}\hat{\theta}(x^n) \tag{11.34}$$

$$= E_{x^n \leftarrow p_\theta^n(x^n)}(\frac{\partial_\theta p_\theta^n(x^n)|_{\theta=0}}{p_\theta^n(x^n)}\hat{\theta}(x^n)) \tag{11.35}$$

$$= E_{x^n \leftarrow p_\theta^n(x^n)}(\partial_\theta \log p_\theta^n(x^n)|_{\theta=0}\hat{\theta}(x^n)) \tag{11.36}$$

One can define an inner product $a \cdot b$ over functions a, b of x^n and apply the Cauchy-Schwarz Inequality over these functions.

$$a \cdot b := E_{x^n \leftarrow p_\theta(x^n)} a(x^n)b(x^n) \tag{11.37}$$

$$(a \cdot b)^2 \leq (a \cdot a)(b \cdot b) \tag{11.38}$$

In our current case, $a(x^n) = \partial_\theta \log p_\theta^n(x^n)|_{\theta=0}$ and $b(x^n) = \hat{\theta}(x^n)$.

$$\begin{aligned}
1^2 &= E_{x^n \leftarrow p_\theta^n(x^n)}(\partial_\theta \log p_\theta^n(x^n)|_{\theta=0}\hat{\theta}(x^n))^2 && (11.39)\\
&= E_{x^n \leftarrow p_\theta^n(x^n)}(a(x^n)b(x^n))^2 && (11.40)\\
&= (a \cdot b)^2 && (11.41)\\
&\leq (a \cdot a)(b \cdot b) && (11.42)\\
&= E_{x^n \leftarrow p_\theta^n(x^n)}a(x^n)E_{x^n \leftarrow p_\theta^n(x^n)}b(x^n) && (11.43)\\
&= E_{x^n \leftarrow p_\theta^n(x^n)}(\partial_\theta \log p_\theta^n(x^n)|_{\theta=0})^2 E_{x^n \leftarrow p_\theta^n(x^n)}(\hat{\theta}(x^n))^2 && (11.44)\\
&= nF\sigma^2(\hat{\theta}) \implies && (11.45)\\
\sigma^2(\hat{\theta}) &\geq \frac{1}{nF} && (11.46)
\end{aligned}$$

□

11.5 Quantum Fisher Information

We will now define the quantum version of Fisher information. In the quantum version of the parameter estimation problem, the parameter θ determines a density matrix ρ_θ from which to sample x, instead of a probability distribution. It is given that $\rho_\theta > 0$ when θ is near $\theta_0 \approx 0$.
We will also now define some new super-operators on matrices.

$$\text{Mult}_\rho(X) = \frac{1}{2}(\rho X + X\rho) \tag{11.47}$$

$$\text{Div}_\rho = \text{Mult}_\rho^{-1} \tag{11.48}$$

$$L_{\rho,\theta} = \text{Div}_{\rho_\theta}(\partial_\theta \rho_\theta) \tag{11.49}$$

$$\partial_\theta \rho_\theta = \text{Mult}_{\rho_\theta}(L_{\rho,\theta}) \tag{11.50}$$

With these super-operators, we can define the quantum Fisher information.

$$F_Q = \text{tr}(\rho L^2) \tag{11.51}$$

Analogously to the classical Fisher information, the quantum Fisher information relates to the second derivative of the quantum relative entropy.

$$F_Q = \partial_\theta^2 D(\rho_\theta || \rho_0)|_{\theta=0} \tag{11.52}$$

We can also relate the quantum Fisher information to the classical Fisher information. To do this, define a set of measurement operators $\{M_x\}_{x\in X}$ with $\sum_{x\in X} M_x = I$. We can calculate the Fisher information of the distribution of the measurement, which is $\rho_\theta(x) = \text{tr}(\rho_\theta M_x)$.

$$\begin{aligned}
F_M &= F(\rho_\theta(x)) && (11.53)\\
&= \sum_{x\in X} \text{tr}(\rho_\theta M_x)(\frac{\partial_\theta \text{tr}(\rho_\theta M_x)}{\text{tr}(\rho_\theta M_x)})^2 && (11.54)\\
&= \sum_{x\in X} \text{tr}(\rho_\theta M_x)(\frac{\text{Re}\{\text{tr}(\rho_\theta L M_x)\}}{\text{tr}(\rho_\theta M_x)})^2 && (11.55)\\
&\leq \sum_{x\in X} \text{tr}(\rho_\theta M_x)(\frac{|\text{tr}(\rho_\theta L M_x)|}{\text{tr}(\rho_\theta M_x)})^2 && (11.56)
\end{aligned}$$

We used the fact that $\partial_\theta \rho_\theta$ is essentially the Hermitian part of $\rho_\theta L_{\rho,\theta}$, so $\partial_\theta \operatorname{tr}(\rho_\theta M_x) = \operatorname{Re}\{\operatorname{tr}(\rho_\theta L M_x)\}$. We can now use the quantum Cauchy-Schwarz Inequality on $|\operatorname{tr}(\rho_\theta L M_x)|$. Specifically, the inequality states that $|\operatorname{tr}(AB)| \leq \sqrt{\operatorname{tr}(A^\dagger A)\operatorname{tr}(B^\dagger B)}$.

$$|\operatorname{tr}(\rho_\theta L M_x)| = |\operatorname{tr}\Big(\sqrt{\rho_\theta} L \sqrt{M_x}\sqrt{M_x}\sqrt{\rho_\theta}\Big)| \tag{11.57}$$

$$\leq \sqrt{\operatorname{tr}(\rho_\theta L M_x L)\operatorname{tr}(\rho_\theta M_x)} \tag{11.58}$$

We can finally substitute this inequality back into F_M.

$$F_M \leq \sum_{x\in X} \operatorname{tr}(\rho_\theta M_x)\Big(\frac{|\operatorname{tr}(\rho_\theta L M_x)|}{\operatorname{tr}(\rho_\theta M_x)}\Big)^2 \tag{11.59}$$

$$\leq \sum_{x\in X} \operatorname{tr}(\rho_\theta L M_x L) \tag{11.60}$$

$$= \operatorname{tr}\big(\rho_\theta L^2\big) \tag{11.61}$$

$$= F_Q \tag{11.62}$$

In summary, the quantum Fisher information of a density matrix is at least the classical Fisher information of distribution of measurement outcomes on that density matrix. The inequality is tight if the measurement is done in the eigenbasis of L.