The problem: want to estimate a parameter encoded in a probability distribution, by looking at a sample from such probability distribution.

Why? Cases in which direct measurement is not possible. E.g. epidemiological models, but also OQS.

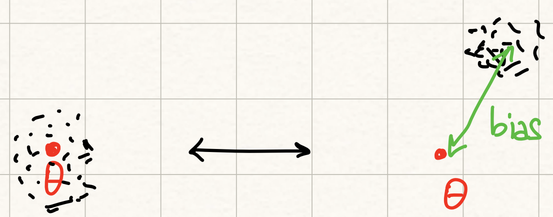
Formal setup

Parameter θ, random variable $X \sim P_\theta(X)$. Take a sample (realisation) of the random variable, x.

Estimator: a function $T: x \mapsto \hat{\theta}$ ("estimate", good guess for θ).

Question: is $\hat{\theta}$ a "good" guess for θ? More precisely, two issues:

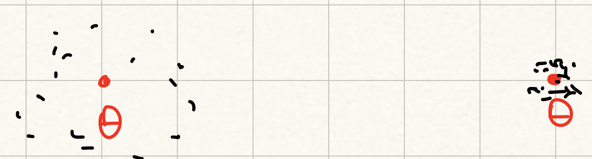
- what is $\mathbb{E}_X[\hat{\theta}(X)]$?



unbiased estimator: $\mathbb{E}_X[\hat{\theta}(X)] = \theta$

Note| This is a frequentist definition, as it requires the existence of a "true θ". Bayesian estimation theory does not directly admit this definition.

- can we say something about the precision of $\hat{\theta}$?



Measure precision by the Mean Squared Error (MSE)

$$\mathbb{E}_X\left[\left(\hat{\theta}(X) - \mathbb{E}_X[\hat{\theta}(X)]\right)^2\right] = \mathrm{Var}_X\left[\hat{\theta}(X)\right] = \mathrm{MSE}(\hat{\theta})$$

for unbiased estimators $\left(\mathbb{E}_X[\hat{\theta}(X)] = \theta\right)$, we have in particular:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}_X\left[(\hat{\theta}(X) - \theta)^2\right]$$

There is a general limit on the precision of an estimator, given by the strength of the parametric encoding in the probability distribution.

$$\mathrm{MSE}_X(\hat{\theta}) \geq \frac{1}{F_\theta(X)} \qquad \text{[Cramér-Rao bound]}$$

Things to note:

(i) it is a lower bound: one can always do worse (larger MSE), but not better (smaller MSE).

(ii) the specific estimator T (or $\hat{\theta}$) does not appear on the RHS $\Rightarrow$ the bound is true for any estimator (equivalently, for the best one).

The denominator:

$$F_\theta(X) = \mathbb{E}_X\left[\left(\underbrace{\frac{\partial}{\partial\theta}}_{\text{how strongly does it depend on } \theta\text{?}} \log \Pr_\theta(X)\right)^2\right] \qquad \text{[Fisher information contained in } X \text{ about } \theta]$$

Equivalent formulation:

$$F_\theta(X) = -\mathbb{E}_X\left[\frac{\partial^2}{\partial\theta^2}\log \Pr_\theta(X)\right]$$

Ex] Prove the equivalence

Proof of the Cramér-Rao bound

Score function $V = \frac{\partial}{\partial\theta}\log \Pr_\theta(X)$

$$\mathbb{E}[V] = \mathbb{E}_X\left[\frac{\partial}{\partial\theta}\log \Pr_\theta(X)\right] = \mathbb{E}_X\left[\frac{1}{\Pr_\theta(X)}\frac{\partial}{\partial\theta}\Pr_\theta(X)\right] =$$

$$= \int dx\, \cancel{\Pr_\theta(x)}^{1}\,\frac{1}{\cancel{\Pr_\theta(x)}_{1}}\frac{\partial}{\partial\theta}\Pr_\theta(x) = \frac{\partial}{\partial\theta}\int dx\,\Pr_\theta(x) = 0$$

Now: $\mathrm{Cov}(V,T) = \mathbb{E}_X\left[T(X)\,\frac{1}{\Pr_\theta(X)}\frac{\partial}{\partial\theta}\Pr_\theta(X)\right] =$

$\overset{\shortparallel}{\mathbb{E}[VT]}$

because $\mathbb{E}[V]=0$

$$= \int dx\, \cancel{\Pr_\theta(x)}^{1}\, T(x)\,\frac{1}{\cancel{\Pr_\theta(x)}_{1}}\frac{\partial}{\partial\theta}\Pr_\theta(x) =$$

$$= \int dx\, T(x)\,\frac{\partial}{\partial\theta}\Pr_\theta(x) = \frac{\partial}{\partial\theta}\int dx\, T(x)\Pr_\theta(x) =$$

$$= \frac{\partial}{\partial\theta}\mathbb{E}\left[T(X)\right]$$

By the Cauchy-Schwartz inequality:

$$\sqrt{\mathrm{Var}[T]\,\mathrm{Var}[V]} \geq |\mathrm{Cov}(V,T)| = \left|\frac{\partial}{\partial\theta}\mathbb{E}[T(X)]\right| \underset{\text{unbiased}}{=} \left|\frac{\partial}{\partial\theta}\theta\right| = 1$$

$$\Rightarrow \quad \mathrm{Var}[T] \geq \frac{1}{\mathrm{Var}[V]} = \frac{1}{\mathbb{E}\left[(V-\underbrace{\mathbb{E}[V]}_{=0})^2\right]} =$$

$$= \frac{1}{\mathbb{E}[V^2]} = \frac{1}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log \Pr_\theta(X)\right)^2\right]} = \frac{1}{F_\theta(X)} . \quad \square$$

Multiple random variables

In general, one does not consider a single random variable, but X as a vector of variables (say of length N).

Simplest case: i.i.d. random variables, $\Pr(X_1 X_2 \dots X_N) = \Pr(X_1)\Pr(X_2)\dots\Pr(X_N)$.

$\hookrightarrow$ the Fisher information is additive. CR-bound:

$$\mathrm{MSE}[T(X_{1:N})] \geq \frac{1}{N F_\theta(X)}$$

this is known as shot-noise scaling or, unfortunately, standard quantum limit.

Saturating the Cramér-Rao bound

Ok, but how to build an estimator? Work through an example:

$$A = \{1,2,3\} \qquad \Pr(1) = \frac{\theta}{2} \qquad \Pr(2) = \frac{1-\theta}{2} \qquad \Pr(3) = \frac{1}{2}$$

Want to estimate θ from a sample like $x_{1:N} = 1\ 2\ 1\ 3\ 2\ 1\ 3\ 3\ 2\ 1 \dots$

Possible estimators:

- sum the string, then ...
- etc.

Optimal estimator? An estimator $T(X)$ is efficient if it saturates the Cramér-Rao bound. In general, we look for asymptotically efficient estimator (the property holds in the $N \to \infty$ limit).

Maximum (log) likelihood estimator

Likelihood function $\quad \ell_x(\theta) = \Pr_\theta(x) \quad$ or $\quad \log \Pr_\theta(x)$ Numerics!

Estimator $\quad \hat{\theta}_{MLE} = \underset{\theta}{\arg\max}\, \ell_x(\theta)$

Idea: the parameter to be chosen is the one that makes the observed sample the most likely.

<u>Theorem</u> (Bernstein-Von Mises, in some variation)
Under appropriate regularity conditions, the maximum likelihood estimator is asymptotically efficient.

Example The maximum likelihood estimator for a binary random variable.

$A = \{0,1\} \qquad \Pr(1) = \theta \qquad \Pr(0) = 1-\theta$

Given a string of length N $x_{1:N}$: a times symbol 1, $N-a$ times symbol 0.

$$\ell_{x_{1:N}}(\theta) = \log \Pr_\theta(x_{1:N}) = \log\left[\theta^a (1-\theta)^{N-a}\right] =$$

$$= a \log\theta + (N-a)\log(1-\theta)$$

Impose maximality: $\frac{\partial}{\partial\theta} \ell_{x_{1:N}}(\theta) \overset{!}{=} 0$

$$\Rightarrow a \frac{1}{\theta} \frac{\partial}{\partial\theta}\theta + (N-a)\frac{1}{1-\theta}\frac{\partial}{\partial\theta}(1-\theta) \overset{!}{=} 0$$

$$\frac{a}{\theta} - (N-a)\frac{1}{1-\theta} \overset{!}{=} 0$$

$$\frac{(1-\theta)a - \theta(N-a)}{\theta(1-\theta)} = 0 \qquad \theta \neq 0,1$$

$$a - \cancel{\theta a} - \theta N + \cancel{\theta a} = 0 \qquad a = \theta N \qquad \theta = \frac{a}{N}$$

That is the average of the obtained values!

Quantum version

So far, in the classical case, the object encoding the parameter is a probability distribution P_θ. In the quantum case, we assume the parameter is encoded in a quantum state.

Problem: there are many ways to go from a state ρ_θ to a probability distribution (many possible measurements).

Therefore: we want to optimise over all the possible measurements.

Idea: define the quantum Fisher information $H_\theta(\rho_\theta)$ as

$$H_\theta(\rho_\theta) = \max_{\{\Pi_x\}} \underbrace{F_\theta}_{\text{classical Fisher information}}(\underbrace{\Pi_x \rho_\theta}_{})$$

classical Fisher information

this is the probability distribution given by the application of the POVM $\{\Pi_x\}$ to the state ρ_θ.

Of course, this maximisation is not very feasible in practice. We find a way to build H_θ analytically.
Remember the definition of the Fisher information

$$F_\theta(X) = \int dx \; \frac{1}{Pr_\theta(x)} \left(\frac{\partial}{\partial\theta} Pr_\theta(x) \right)^2 = (*)$$

but now, by the Born rule: $Pr_\theta(x) = \mathrm{Tr}[\Pi_x \rho_\theta]$ for a given POVM $\{\Pi_x\}$, hence

$$(*) = \int dx \; \frac{1}{\mathrm{Tr}[\Pi_x \rho_\theta]} \left(\frac{\partial}{\partial\theta} \mathrm{Tr}[\Pi_x \rho_\theta] \right)^2 =$$

$$= \int dx \; \frac{1}{\mathrm{Tr}[\Pi_x \rho_\theta]} \left(\mathrm{Tr}\left[\Pi_x \frac{\partial}{\partial\theta} \rho_\theta \right] \right)^2$$

We express the derivative of the state in terms of the action of a Hermitian operator L_θ (known as "symmetric logarithmic derivative"):

$$\frac{\partial \rho_\theta}{\partial\theta} = \frac{L_\theta \rho_\theta + \rho_\theta L_\theta}{2} = \quad [\text{Lyapunov equation}]$$

$$= \mathrm{Re}[L_\theta \rho_\theta]$$

$$\Rightarrow \mathrm{Tr}\left[\Pi_x \frac{\partial}{\partial\theta} \rho_\theta\right] = \mathrm{Re}\left[\mathrm{Tr}\left[\Pi_x L_\theta \rho_\theta\right]\right] = \mathrm{Re}\left[\mathrm{Tr}\left[\rho_\theta \Pi_x L_\theta\right]\right]$$

Hence, the classical Fisher information is

$$F_\theta(\Pi_x \rho_\theta) = \int dx \; \frac{1}{\mathrm{Tr}[\Pi_x \rho_\theta]} \left(\mathrm{Re}\left[\mathrm{Tr}\left[\rho_\theta \Pi_x L_\theta\right]\right] \right)^2 \leq (1)$$

Now, we want to find a maximisation in terms of something that does not depend on the choice of the specific POVM $\{\Pi_x\}$.

$$\leq \int dx \left| \frac{\mathrm{Tr}[\rho_\theta \Pi_x L_\theta]}{\sqrt{\mathrm{Tr}[\rho_\theta \Pi_x]}} \right|^2 = \int dx \left| \mathrm{Tr}\left[\frac{\sqrt{\rho_\theta}\sqrt{\Pi_x}}{\sqrt{\mathrm{Tr}[\rho_\theta \Pi_x]}} \sqrt{\Pi_x} \quad \sqrt{\rho_\theta} \right] \right|^2 =$$

Using the Cauchy-Schwartz inequality:

$$= \int dx \left| \left\langle \frac{\sqrt{\rho_\theta}\sqrt{\Pi_x}}{\sqrt{\mathrm{Tr}[\rho_\theta \Pi_x]}} , \sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta} \right\rangle \right|^2 \leq (2)$$

$$\leq \int dx \left\| \frac{\sqrt{\rho_\theta}\sqrt{\Pi_x}}{\mathrm{Tr}[\rho_\theta \Pi_x]} \right\|^2 \left\| \sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta} \right\|^2 =$$

$$= \int dx \frac{\overbrace{\cancel{\mathrm{Tr}[\rho_\theta \Pi_x]}}^{1}}{\underbrace{\cancel{\mathrm{Tr}[\rho_\theta \Pi_x]}}_{1}} \mathrm{Tr}\left[\sqrt{\rho_\theta} L_\theta \sqrt{\Pi_x}\sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta} \right] =$$

$$= \int dx \, \mathrm{Tr}\left[\Pi_x L_\theta \rho_\theta L_\theta \right] =$$

Now use the normalisation of the POVM $\{\Pi_x\}$, $\int dx \, \Pi_x = \mathbb{1}$, hence

$$= \mathrm{Tr}\left[\rho_\theta L_\theta^2 \right]$$

This does not depend on the POVM $\{\Pi_x\}$! Therefore:

$$F_\theta(\Pi_x \rho_\theta) \leq \mathrm{Tr}\left[\rho_\theta L_\theta^2 \right] = H_\theta(\rho_\theta) \qquad \text{[Quantum Fisher information]}$$

Follows the quantum Cramér–Rao bound:

$$\mathrm{MSE}(T, \rho_\theta) \geq \frac{1}{H_\theta(\rho_\theta)}$$

Optimal POVMs saturating the Cramér-Rao bound

We have to choose an optimal POVM $\{\Pi_x\}$, that saturates both inequalities (1) and (2).
The inequality (1) is saturated if $\mathrm{Tr}[\rho_\theta \Pi_x L_\theta]$ is real for all θ.
The inequality (2) is saturated when

$$\frac{\sqrt{\Pi_x}\sqrt{\rho_\theta}}{\mathrm{Tr}[\rho_\theta \Pi_x]} = \frac{\sqrt{\Pi_x} L_\theta \sqrt{\rho_\theta}}{\mathrm{Tr}[\rho_\theta \Pi_x L_\theta]}$$

(i.e., the two vectors in the Cauchy–Schwartz are parallel)

This is satisfied if and only if (more or less) $\{\Pi_x\}$ is the set of projectors over eigenstates of L_θ. The optimal POVM is L_θ.

Technical note The optimal POVM L_θ yields the maximal Fisher information, coinciding with the quantum Fisher information.
However, this says nothing on the optimal estimator, i.e. the optimal function of the eigenvalues of L_θ. One can apply maximum likelihood.

So:
- we can find the maximal amount of parameter knowledge extractable from a quantum state.
- we can write the optimal POVM.

All of this requires to be able to compute the symmetric logarithmic derivative L_θ, which was implicitly defined by the Lyapounov equation

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

It is solved (given w.o. proof) by

$$L_\theta = 2 \sum_{nm} \frac{\langle \psi_m | \partial_\theta \rho_\theta | \psi_n \rangle}{p_n + p_m} |\psi_m\rangle\langle\psi_n| \quad \text{for terms w. } p_n + p_m \neq 0.$$

where $\rho_\theta = \sum_n p_n |\psi_n\rangle\langle\psi_n|$ (eigenbasis decomposition)

Note that, in general, both the eigenvalues $\{p_n\}$ and the eigenstates $\{|\psi_n\rangle\}$ depend on the value of the parameter θ.

<u>Ex</u> The derivative of the state ρ_θ can be written as

$$\frac{\partial}{\partial\theta} \rho_\theta = \sum_n \left(\frac{\partial p_n}{\partial\theta} |\psi_n\rangle\langle\psi_n| + p_n |\partial_\theta \psi_n\rangle\langle\psi_n| + p_n |\psi_n\rangle\langle\partial_\theta \psi_n| \right)$$

Note that $\langle\psi_m|\psi_n\rangle = \delta_{mn} \Rightarrow \partial_\theta \left(\langle\psi_m|\psi_n\rangle\right) = \langle\partial_\theta \psi_m|\psi_n\rangle + \langle\psi_m|\partial_\theta\psi_n\rangle = 0$

$$\Rightarrow \text{Re}\, \langle\partial\psi_n|\psi_m\rangle = 0 \,, \quad \langle\partial_\theta\psi_m|\psi_n\rangle = -\langle\psi_m|\partial_\theta\psi_n\rangle$$

Applying the solution of the Lyapounov equation for the symmetric logarithmic derivative L_θ, and the definition of the quantum Fisher information, prove that the QFI can be decomposed in terms of a classical contribution by the distribution of the eigenvalues $\{p_n\}$, and a "truly quantum" one in terms of derivatives of the basis $\{|\psi_n\rangle\}$.

$$L_\theta = 2\sum_{mn} \frac{\langle\psi_m|\partial_\theta\rho_\theta|\psi_n\rangle}{\rho_m+\rho_n}|\psi_m\rangle\langle\psi_n| =$$

$$= 2\sum_{mn}\frac{1}{\rho_m+\rho_n}\langle\psi_m|\left(\sum_k \frac{\partial\rho_k}{\partial\theta}|\psi_k\rangle\langle\psi_k| + \rho_k|\partial_\theta\psi_k\rangle\langle\psi_k| + \rho_k|\psi_k\rangle\langle\partial_\theta\psi_k|\right)|\psi_n\rangle|\psi_m\rangle\langle\psi_n| =$$

$$= 2\sum_{mnk}\frac{|\psi_m\rangle\langle\psi_n|}{\rho_m+\rho_n}\left(\frac{\partial\rho_k}{\partial\theta}\overbrace{\langle\psi_m|\psi_k\rangle}^{\delta_{mk}}\overbrace{\langle\psi_k|\psi_n\rangle}^{\delta_{kn}} + \rho_k\langle\psi_m|\partial_\theta\psi_k\rangle\overbrace{\langle\psi_k|\psi_n\rangle}^{\delta_{kn}} + \rho_k\overbrace{\langle\psi_m|\psi_k\rangle}^{\delta_{km}}\langle\partial_\theta\psi_k|\psi_n\rangle\right)$$

$$= 2\sum_k \frac{\partial\rho_k}{\partial\theta}\frac{1}{2\rho_k}|\psi_k\rangle\langle\psi_k| + 2\sum_{mn}\frac{|\psi_m\rangle\langle\psi_n|}{\rho_m+\rho_n}\left(\rho_n\langle\psi_m|\partial_\theta\psi_n\rangle + \rho_m\langle\partial_\theta\psi_m|\psi_n\rangle\right) =$$

$$= \sum_k \frac{\partial\rho_k}{\partial\theta}\frac{1}{\rho_k}|\psi_k\rangle\langle\psi_k| + 2\sum_{mn}\frac{\rho_n-\rho_m}{\rho_m+\rho_n}\langle\psi_m|\partial_\theta\psi_n\rangle|\psi_m\rangle\langle\psi_n|$$

$$\Rightarrow H_\theta(\rho_\theta) = \sum_k\left(\frac{\partial\rho_k}{\partial\theta}\frac{1}{\rho_k}\right)^2 \overbrace{\mathrm{Tr}\left[|\psi_k\rangle\langle\psi_k|\rho_\theta\right]}^{\rho_k} + \dots =$$

$$= \underbrace{\sum_k \frac{1}{\rho_k}\left(\frac{\partial\rho_k}{\partial\theta}\right)^2}_{\text{classical term}} + \underbrace{\dots}_{\text{quantum correction}}$$

So far, everything is a bit abstract. Now we specialise the results to some cases of physical interest.

Unitary encoding

This is the most typical case in quantum metrology: send in a probe state into a unitary channel w. an unknown phase.

$$\rho_0 \longrightarrow \boxed{U_\theta} \longrightarrow \rho_\theta \qquad U_\theta = e^{-i\theta H} \qquad H = H^\dagger \text{ (some Hamiltonian)}$$

$$\Rightarrow \rho_\theta = U_\theta \rho_0 U_\theta^\dagger = e^{-i\theta H} \rho_0 e^{i\theta H}$$

Derivative of ρ_θ for the symmetric logarithmic derivative.

$$\frac{\partial}{\partial\theta}\rho_\theta = \left(\frac{\partial}{\partial\theta} U_\theta\right)\rho_0 U_\theta^\dagger + U_\theta \rho_0 \left(\frac{\partial}{\partial\theta} U_\theta^\dagger\right) =$$

$$= -iH U_\theta \rho_0 U_\theta^\dagger + i U_\theta \rho_0 H U_\theta^\dagger = \qquad [H, U_\theta] = 0$$

$$= -i U_\theta H \rho_0 U_\theta^\dagger + i U_\theta \rho_0 H U_\theta^\dagger =$$

$$= i U_\theta [\rho_0, H] U_\theta^\dagger$$

$$L_\theta = 2 \sum_{nm} \frac{\langle \psi_m | \partial_\theta \rho_\theta | \psi_n \rangle}{p_n + p_m} |\psi_m\rangle\langle\psi_n|$$

where $\{|\psi_m\rangle\}$ eigenstates of ρ_θ with eigenvalues $\{p_m\}$. But it is convenient to express everything in terms of eigenstates of ρ_0: $\{|\varphi_n\rangle\}$, s.t.

$$|\varphi_n\rangle = U_\theta^\dagger |\psi_n\rangle, \quad |\psi_n\rangle = U_\theta |\varphi_n\rangle$$

The eigenvalues are preserved by the unitary evolution.

$$\Rightarrow L_\theta = 2 \sum_{mn} \frac{\langle \varphi_m | \overbrace{U_\theta^\dagger \, i \, U_\theta}^{\mathbb{1}} [\rho_0, H] \overbrace{U_\theta^\dagger U_\theta}^{\mathbb{1}} | \varphi_n \rangle}{p_n + p_m} U_\theta |\varphi_m\rangle\langle\varphi_n| U_\theta^\dagger =$$

$$= 2i \sum_{mn} \frac{\langle \varphi_m | [\rho_0, H] | \varphi_n \rangle}{\rho_n + \rho_m} \underbrace{U_\theta} | \varphi_m \rangle\langle \varphi_n | \underbrace{U_\theta^\dagger} = U_\theta L_0 U_\theta^\dagger$$

the only dependence on θ is here

$$\Rightarrow L_0 = 2i \sum_{mn} \frac{\langle \varphi_m | [\rho_0, H] | \varphi_n \rangle}{\rho_n + \rho_m} | \varphi_m \rangle\langle \varphi_n | = \quad , \text{ but } \langle \varphi_m | \rho_0 = \rho_m \, , \, \rho_0 | \varphi_n \rangle = \rho_n$$

$$= 2i \sum_{mn} \frac{\rho_m - \rho_n}{\rho_n + \rho_m} \langle \varphi_m | H | \varphi_n \rangle | \varphi_m \rangle\langle \varphi_n |$$

$$\Rightarrow H_\theta(\rho_\theta) = \mathrm{Tr}\left[\rho_\theta L_\theta^2\right] = \mathrm{Tr}\left[\partial_\theta \rho_\theta L_\theta\right] = \mathrm{Tr}\left[\partial_\theta \rho_\theta \, U_\theta L_0 U_\theta^\dagger\right] =$$

$$= \mathrm{Tr}\left[i \, U_\theta [\rho_0, H] \underbrace{U_\theta^\dagger U_\theta}_{\mathbb{1}} L_0 U_\theta^\dagger\right] = i \, \mathrm{Tr}\left[[\rho_0, H] L_0\right] =$$

$$= -i \, \mathrm{Tr}\left[H [\rho_0, L_0]\right] =$$

$$= 2 \sum_{m,n} \frac{\rho_m - \rho_n}{\rho_n + \rho_m} \langle \varphi_m | H | \varphi_n \rangle \, \mathrm{Tr}\left[H \left[\rho_0, | \varphi_m \rangle\langle \varphi_n |\right]\right] =$$

$$= 2 \sum_{mn} \frac{\rho_m - \rho_n}{\rho_n + \rho_m} \langle \varphi_m | H | \varphi_n \rangle \left(\langle \varphi_n | H \rho_0 | \varphi_m \rangle - \langle \varphi_n | \rho_0 H | \varphi_m \rangle \right) =$$

$$= 2 \sum_{mn} \frac{\rho_m - \rho_n}{\rho_n + \rho_m} \langle \varphi_m | H | \varphi_n \rangle \underline{\underline{\langle \varphi_n | [H, \rho_0] | \varphi_m \rangle}}$$

The information about θ is collected if $[H, \rho_0] \neq 0$: do not send eigenstate of H!