Quantum Fluctuation-Response Inequality and Its Application in Quantum Hypothesis Testing

Yan Wang
Department of Statistics, Iowa State University, Ames, Iowa 50011, USA

We uncover the quantum fluctuation-response inequality, which, in the most general setting, establishes a bound for the mean difference of an observable at two different quantum states, in terms of the quantum relative entropy. When the spectrum of the observable is bounded, the sub-Gaussian property is used to further our result by explicitly linking the bound with the sub-Gaussian norm of the observable, based on which we derive a novel bound for the sum of statistical errors in quantum hypothesis testing. This error bound holds nonasymptotically and is stronger and more informative than that based on quantum Pinsker’s inequality. We also show the versatility of our results by their applications in problems like thermodynamic inference and speed limit.

Introduction.— Quantifying the mean difference of a quantity at two states is of great interest in both classical and quantum settings, and in various fields from physics to machine learning and statistics. For example, in physics, when some system parameter is perturbed from θ₀ to θ = θ₀ + δθ, the system property will be changed accordingly. When δθ is small, the quantitative difference between \( \text{Tr}(O_0) \) and \( \text{Tr}(O) \) is captured by the celebrated linear response theory \[1\], where \( O \) is some observable of interest and \( \Phi \) the test function. This expression states that the mean difference (classical or quantum) is small, \( \delta \theta \), \( \Phi \) is some observables of interest and \( \rho_0 \) and \( \rho \) are the corresponding density operators of the original and the perturbed system, respectively. In machine learning, one central task is to quantify the difference between the empirical risk and the expected risk \[2\]. That is, for a training set of independent and identically distributed data \( \{ (X_i, Y_i) \}_{i=1}^n \) and an algorithm \( A \) trained on it, it is important to bound the difference between the out-of-sample risk \( \mathbb{E}_{Z \sim P_2} L(Y, A(X)) \) and the in-sample risk \( \mathbb{E}_{Z \sim P_1} L(Y, A(X)) \), where \( L \) is the loss function, \( P_2 \) is the underlying data-generating distribution, and \( P_1 = \frac{1}{n} \sum_{i=1}^n \delta_{Z_i} \) is the empirical distribution of the training set with \( \delta_{Z_i} = \delta(Z - z) \) being the Dirac delta function.

In statistics, for a test function \( \Phi, \mathbb{E}_{H_0} \Phi = \alpha \) is the type I error (false positive) rate, and \( 1 - \mathbb{E}_{H_1} \Phi = \beta \) gives the type II error (false negative) rate, where \( H_0 \) and \( H_1 \) denote the hypotheses under the null hypothesis \( H_0 \) and the alternative hypothesis \( H_1 \), respectively \[3\]. Note that \( \mathbb{E}_{H_0} \Phi - \mathbb{E}_{H_1} \Phi \) quantifies the sum of error rates \( \alpha + \beta \) up to a constant 1, hence quantifying this difference is important in analyzing and optimizing the test function \( \Phi \) \[4\]. On a high level, people seek to establish in many cases that

\[
|\mathbb{E}_0 O - \mathbb{E}_1 O| \leq f(\text{system, operation}).
\]

This expression states that the mean difference (classical or quantum) of \( O \) at two different states is bounded by a function \( f \) of the properties of the original and perturbed systems and the way we manipulate the original system. In linear response, such a mean difference is equal to \( \chi \delta \theta \) with \( \delta \theta \) showing to what extent the system is modified experimentally and the associated susceptibility \( \chi \) reflecting the thermodynamic property of the original system \[1\]. In machine learning and statistics, the results are usually in the form of inequality, in terms of some distance measure between probabilities and the complexity measure of a given training algorithm or test function \[2, 4, 6\].

Recently, an elegant framework for analyzing the mean difference of a generic quantity between two arbitrary nonequilibrium states is established, i.e., the fluctuation-response inequality (FRI) \[7\]. However, this idea is not only restricted to the classical case; it is also of great interest in the quantum realm. Moreover, its potential applicability in science can go way beyond the scope of statistical physics. The aim of this work is two-fold. First, we establish a quantum version of the FRI, which characterizes the mean difference of a generic observable at two arbitrary quantum states. Second, by taking advantage of the sub-Gaussian property associated with any measurement plan, we provide an operationally useful bound of error rates for quantum hypothesis testing \[8\], which is stronger and thus more informative than that from quantum Pinsker’s inequality. We also discuss possible applications of our results in such topics like thermodynamic inference \[9\] and speed limit \[10\]. As a general approach, our results are expected to find applications in various fields of science.

Quantum fluctuation-response inequality.— The major difficulty in generalizing the classical fluctuation-response theory to the quantum regime lies in the fact that operators in general do not commute. We tackle this problem by carefully constructing useful operators from the most general setting. Given two arbitrary density operators \( \gamma_0 \) and \( \gamma_1 \) that describe a system at two different states, we assume \( \gamma_0 \) is full-rank in the Hilbert space of interest, which describes the reference system, and \( \gamma_1 \) the “perturbed” system. For an arbitrary Hermitian operator \( O \), we want to bound \( |\text{Tr}(O \gamma_0) - \text{Tr}(O \gamma_1)| \) where the traces are supposed to be well defined. First, let \( O_c \equiv O - \text{Tr}(O \gamma_0) I \) be the “centered” version of \( O \), where \( I \) is the identity operator. It is easy to see that \( \text{Tr}(O_0 \gamma_0) = 0 \). Then, for \( s > 0 \), we define:

\[
B = \xi s O_c + \ln \gamma_0, \quad (1)
\]

\[
A = \ln \gamma_1 + \lambda I, \quad (2)
\]

where \( \lambda \equiv \text{Tr}[\gamma_1 (B - \ln \gamma_1)], \xi = \text{sgn}(\text{Tr}(O \gamma_1) - \text{Tr}(O \gamma_0)) \) and \( \text{sgn}(\cdot) \) is the sign function. One can verify that

\[
\text{Tr}[(B - A) \gamma_1] = 0. \quad (3)
\]

To proceed, let us recall Klein’s inequality, which is a quantum version of the property of convex functions, and by which we obtain \( \text{Tr}(e^B) \geq \text{Tr}(e^A) + \text{Tr}(e^A(B - A)) \). Inserting Eq.
\begin{align}
\text{(4)} \quad \ln \text{Tr}(e^{sO_c+\ln \gamma_0}) & \geq \xi_s [\text{Tr}(\gamma_0) - \text{Tr}(\gamma_0 O)] = s[\text{Tr}(\gamma_0 O) - \text{Tr}(\gamma_0 || \gamma_0)], \\
& = s[\text{Tr}(\gamma_0 O) - \text{Tr}(\gamma_0 O)] - S(\gamma_0 || \gamma_0),
\end{align}

where the second inequality is due to the Golden-Thompson inequality [13, 14]. This is the first main result of this work, which may be termed as the quantum fluctuation-response inequality (QFRI). Compared to its classical counterpart \cite{7}, one can see that formally they are identical. However, despite this formal similarity, it is worth stressing that there is an important difference between (4) and its classical counterpart. In the latter, the probabilities can be calculated exactly for all \( s > 0 \). Hence, take the infimum with respect to \( s \), we obtain

\[ |\text{Tr}(\gamma_0 O) - \text{Tr}(\gamma_0 O)| \leq \inf_{s > 0} \frac{1}{s} \left[ \ln \text{Tr}(e^{sO_c+\ln \gamma_0}) + S(\gamma_0 || \gamma_0) \right], \]

where

\( \sigma^2 \) such that

\[ \mathbb{E}_{X \sim P} \exp[t(X - \mathbb{E}_{X \sim P} X)] \leq e^{\sigma^2 t^2/2} \]

for all \( t \in \mathbb{R} \). The square root of the infimum of all such \( \sigma^2 \) is defined to be the sub-Gaussian norm of \( X \), denoted \( \sigma_{OP} \). Many distributions are sub-Gaussian, including the Gaussian itself. In particular, all bounded random variables are sub-Gaussian. This fact is especially relevant to the discussion in quantum information science where usually a finite dimensional Hilbert space is concerned, and physical quantities of interest are typically bounded \cite{8, 12}. Hence, the sub-Gaussian property can be extremely helpful in such cases.

To take advantage of the sub-Gaussianity, we use an orthonormal basis of the Hilbert space in question that diagonalizes \( O_c \), such that

\[ \text{Tr}(\gamma_0 e^{sO_c}) = \sum_i \langle i | \gamma_0 | i \rangle e^{s \xi_i} \],

where \( O_c = \sum_i \alpha_i | i \rangle \langle i | \) is the eigendecomposition of \( O_c \). Given \( \gamma_0 = O_c \) is a density operator, it is legitimate to interpret \( P \equiv \{ p_i = \langle i | \gamma_0 | i \rangle \} \) as a probability distribution with \( p_i \in [0, 1] \) for all \( i \). We may also define \( O_c \) as the corresponding random variable that follows the distribution \( P \), which equals \( c_i \) with probability \( p_i \). This way, we rewrite the trace as an expectation:

\[ \text{Tr}(\gamma_0 e^{sO_c}) = \mathbb{E}_{O_c \sim P} (e^{sO_c}). \]

Note \( \mathbb{E}_{O_c \sim P} O_c = 0 \). Hence if the eigenvalues of \( O_c \) are bounded, then \( O_c \) is sub-Gaussian, and its norm is denoted \( \sigma_{OP} \). By (5), we obtain

\[ \text{Tr}(\gamma_0 e^{sO_c}) \leq \exp[\sigma_{OP}^2 (s^2/2)] = \exp(\sigma_{OP}^2 s^2/2). \]

Inserting this to (4) yields

\[ |\text{Tr}(\gamma_0 O) - \text{Tr}(\gamma_0 O)| \leq \inf_{s > 0} \frac{1}{s} \left[ \ln e^{\sigma_{OP}^2 s^2/2} + S(\gamma_0 || \gamma_0) \right] = \sigma_{OP} \sqrt{2S(\gamma_0 || \gamma_0)}, \]

where the infimum is achieved at \( s = \sqrt{2S(\gamma_0 || \gamma_0)}/\sigma_{OP} \). This sub-Gaussian QFRI is our second main result, which is more relevant, than (4), to cases where operators of interest have a bounded spectrum. One practical question is how to obtain \( \sigma_{OP} \), which might also be defined to be the sub-Gaussian norm of \( O \) with respect to \( \gamma_0 \). It is known that if \( X \) is bounded in \([a, b]\), then \( \sigma_{XP} \leq (b - a)/2 \) holds universally irrespective of \( P \) \cite{16, 17}. However, given \( P \), one can numerically calculate \( \sigma_{XP} \) in a principled way or find an informative upper bound, which will be addressed below.

By (6), new bounds of possible physical interest can be obtained. For example, let us pick \( O \approx \mathcal{H} \) with \( \mathcal{H} \) the Hamiltonian. Without loss of generality, the eigenvalues of \( \mathcal{H} \) are supposed to be bounded in \([0, H_{\text{max}}] \) for a finite system. Then (6) implies that \( |U_1 - U_0| \leq \sigma_{\mathcal{H}P} \sqrt{2S(\gamma_0 || \gamma_0)} \) with \( \sigma_{\mathcal{H}P} \leq H_{\text{max}}/2 \) being the sub-Gaussian norm of \( \mathcal{H} \) with respect to \( \gamma_0 \) and \( U_{0, 1} = \text{Tr}(\gamma_0 \mathcal{H}) \) being the corresponding “internal energies” of the system at \( \rho_{0, 1} \). Hence the mean energy difference at different states can be bounded in terms of the quantum relative entropy between the states. If we further let \( \gamma_0 = e^{-\mathcal{H}/T} / \text{Tr}(e^{-\mathcal{H}/T}) \) be a thermal state at temperature \( T \), and \( \gamma_1 \) can still be an arbitrary nonequilibrium state, then we obtain

\[ |U_1 - U_0| \leq \sigma_{\mathcal{H}P} \sqrt{2(F_1 - F_0)/T} \]

with some \( \mathcal{H} \) into the expression, and noting Eq. (3) and the fact that the commutator \([\ln \gamma, I] = 0\), we then have \cite{11}

\[ \text{Tr}(e^B) \geq \text{Tr} \left( e^{\ln \gamma + \lambda I} + \text{Tr} \left( e^{\ln \gamma + \lambda I} (B - A) \right) \right) = \exp(\text{Tr} [\gamma (B - \ln \gamma)]). \]
$F_{0,1} \equiv U_{0,1} - TS_{0,1}$ being “Helmholtz free energies” and $S_{0,1} = -\text{Tr}(\gamma_0 \ln \gamma_{0,1})$ being entropies at states $\rho_{0,1}$, respectively. Note $T$ is fixed in this case, and at equilibrium $F_0$ is minimized, hence $F_1$ is greater than $F_0$ in general.

We believe \cite{4} and \cite{5} as general bounds can lead to physically insightful results when applied to specific problems. In the following, we focus on the application of \cite{6} in quantum hypothesis testing where $O_c$ is constructed from a measurement plan.

Quantum hypothesis testing.— Classical statistics deals with the discrimination of two probabilities, while quantum hypothesis testing aims to discriminate two quantum states $\rho_0$ and $\rho_1$. From the classical to the quantum setting, some similar ideas are shared in constructing optimal tests, and some classical results have their quantum counterparts with almost the same form \cite{20,22}. However, one fundamental difference is that in the latter case, measurement also determines the outcome distribution in addition to the quantum state, and thus is an additional factor to consider which has no classical counterpart. Also, in the quantum case, one can incorporate the test step into the measurement step to compactly design a positive operator-valued measurement (POVM) $M = \{M_0, M_1\}$, where $M_{0,1}$ are positive semidefinite operators with $M_0 + M_1 = I$. $M_0$ corresponds to the outcome equal to 0, meaning $\rho_0$ is accepted; similarly, $M_1$ corresponds to the outcome 1 and accepts $\rho_1$. Given $n$ copies of the system, the error rates are calculated as $\alpha = \text{Tr}(\rho_0^{\otimes n} M_1)$ and $\beta = \text{Tr}(\rho_1^{\otimes n} M_0)$. Similar to the classical case, it is known that $\alpha + \beta \geq 1 - \frac{1}{2} \parallel \rho_0^{\otimes n} - \rho_1^{\otimes n} \parallel_1$, where $\parallel \cdot \parallel_1$ denotes the trace norm, but the calculation of $\parallel \rho_0^{\otimes n} - \rho_1^{\otimes n} \parallel_1$ can be prohibitive as the dimension of $\rho^{\otimes n}$ increases exponentially fast. Thanks to quantum Pinsker’s inequality \cite{23}, an alternative upper bound can be found as

$$\alpha + \beta \geq 1 - \frac{1}{2} \sqrt{2nS(\rho_1|\rho_0)},$$

which may be practically useful when $S(\rho_1|\rho_0)$ is small. This result is universal in the sense that it holds for any POVM used to perform the hypothesis testing. Moreover, the constant coefficient $\frac{1}{2}$ is sharp.

That said, we will show by using \cite{4}, a stronger and more informative bound than (7) can be achieved, which takes the measurement plan into account. This is important as we do want to know an error bound associated with a given measurement plan, while \cite{7} only provides a bound for the optimal POVM $M_{\text{opt}}$. Experimentally, it is not always possible to implement $M_{\text{opt}}$ due to technical difficulties or simply because $M_{\text{opt}}$ cannot be easily found theoretically. Now, suppose $M$ is the measurement plan designed for $n$ copies of the system. To establish a bound associated with $M$, we set $O = M_1$ and $\gamma_0 = \rho_0^{\otimes n}$ in \cite{6}. Since $\text{Tr}(\gamma_0 M_1) = \alpha$ and $\text{Tr}(\gamma_1 M_1) = \text{Tr}(\gamma_1 (I - M_0)) = 1 - \beta$, we have

$$\alpha + \beta \geq 1 - \sigma_0 \sqrt{2nS(\rho_1|\rho_0)},$$

where $\sigma_0$ is the sub-Gaussian norm of some measurement-dependent random variable detailed below. For practically nontrivial cases with $\alpha < 0.5$, $\sigma_0 \leq \sqrt{\frac{\alpha - 0.5}{\ln(1/(1-\alpha))}} < \frac{1}{2}$. This is the third main result of this work. Before we show how to obtain or bound $\sigma_0$, several remarks are in order. First, our bound evidently consists of two parts. $S(\rho_1|\rho_0)$ describes the distance between two quantum states, while $\sigma_0$ characterizes the performance of the very POVM we use to implement the test. This explains why our bound is more informative than \cite{7}, which bounds the best performance of all possible measurement plans, and thus conservative. Using \cite{6}, we only need to compute $\alpha = \text{Tr}(\gamma_0 M_1)$ once and a sensible error bound can be obtained. The explicit dependence on $\alpha$ is perhaps the most distinctive feature of our result. Second, $\sigma_0 < \frac{1}{2}$ holds in all nontrivial cases where $\alpha \neq 0.5$ (Note a random guess leads to $\alpha = 0.5$.) Hence \cite{6} is strictly stronger than \cite{7}, especially when $\alpha \ll 0.5$. Third, this result is a nonasymptotic one that holds for any finite $n$. This is different than asymptotic results like the quantum Chernoff bound and quantum Stein’s lemma \cite{20,22}, which only hold as $n \to \infty$. Fourth, by swapping $H_0$ and $H_1$, there is a natural twin of \cite{6} that $\beta + \alpha \geq 1 - \sigma_1 \sqrt{2nS(\rho_0|\rho_1)}$. In general, $S(\rho_1|\rho_0) \neq S(\rho_0|\rho_1)$ and $\sigma_0 \neq \sigma_1$, thus these two inequalities are not the same. Nonetheless, \cite{6} is perhaps more useful practically since it is usually desirable to control $\alpha$ at a low level, resulting in a low $\sigma_0$. Fifth, from \cite{6}, there is also an upper bound $\alpha + \beta \leq 1 + \sigma_0 \sqrt{2nS(\rho_1|\rho_0)}$. However, this is not quite useful since a trivial test by always accepting $\rho_0$ or $\rho_1$ leads to $\alpha + \beta = 1$. It might be more informative in the Bayesian setting, though \cite{11}.

Calculating $\sigma_0$— Now let us turn to $\sigma_0$. We first expand $M_1$ as $M_1 = \sum m \mu_m |m\rangle \langle m|$, where $\{ |m\rangle \}$ is an orthonormal basis in the whole Hilbert space. Since $0 \leq M_1 \leq I$, we have $\mu_m \in [0,1]$ for all $m$. Then $\alpha = \text{Tr}(\gamma_0 M_1) = \sum m \mu_m p_m$ with $p_m = \langle m|\gamma_0|m\rangle$. Similarly, we have $\text{Tr}(\gamma_1 M_1) = \sum m \mu_m p_m e^{-\xi_0}$, and define a random variable $M$ that follows probability $P_M$ as follows: $M = 0$ with probability $1 - q$ and $M = \mu_m$ with probability $p_m$ for all $m$ with $\mu_m > 0$. Clearly, $M$ is bounded in $[0,1]$, hence it is straightforward to know that $M$ is sub-Gaussian with norm $\sigma_M P_M \equiv \sigma_0 \leq 1/2$. Noting $E_{M \sim P_M} M = \alpha$, we define $K(t) \equiv \ln E_{M \sim P_M} e^{t(M - \alpha)}$ for $t \in \mathbb{R}$, then by definition \cite{5}, one can calculate $\sigma_0$ by solving a set of equations that $K(t) = \frac{1}{2} \sigma^2 t^2$ and $dK(t)/dt = \sigma^2 t$ \cite{24}. The minimal solution of $\sigma$ (if there exist more than one solution) gives $\sigma_0$. The underlying idea is that by decreasing $\sigma$ to the critical $\sigma_0$, $\sigma^2 t^2/2$ must intersect $K(t)$ at some $t^*$ (the first equation), and this is achieved when they are tangent to each other at $t^*$ (the second equation). Note the trivial case $t = 0$ should be excluded as $K(0) = K(0) = 0$ irrespective of $\sigma$.

When $M_1$ is a projector, we can find $\sigma_0$ explicitly. In this case, the eigenvalues of $M_1$ are in $\{0,1\}$, hence correspondingly $M$ is a Bernoulli random variable with mean $\alpha = \sum m \mu_m > 0 \mu_m p_m = \sum m \mu_m = 1 p_m$, and thus $\text{Tr}(\gamma_0 e^{\xi_0 M_1}) = (1 - \alpha) e^{-\xi_0} + \alpha e^{\xi_0 (1-\alpha)} \leq e^{\sigma^2 s^2/2}$. For
\( \alpha \in (0, 1) \) and \( \alpha \neq 0.5 \), it is known that \( \sigma_0 = \sqrt{\frac{\alpha - 0.5}{\ln(\alpha/(1-\alpha))}} \) \cite{13}, which is strictly less than 0.5. Only when \( \alpha = 0.5 \), we have \( \sigma_0 = 0.5 \), and this is the only situation that \( \sigma_0 \) reduces to \( \sigma_0 \). It is worth noting that when \( M_1 \) is a projector, \( \sigma_0 \) surprisingly only depends on the false positive rate \( \alpha \) without resorting to the details of the measurement plan, because the information of the eigenvalues of \( M_1 \) has been implicitly encoded in \( \alpha \). Hence if some POVM is designed to control the false positive rate at a low level \( \alpha \), then we immediately know the resulting \( \sigma_0 \) and the corresponding lower bound for the sum of error rates. In particular, in the extreme case that \( \alpha \ll 0.5 \), we approximately have \( \sigma_0 \approx \sqrt{\frac{1}{2m}} \), which indicates in such an extreme situation that \( \alpha \) is greatly suppressed, our bound is nontrivial (greater than 0) as long as \( nS(\rho_1\|\rho_0) \leq \ln(\frac{1}{\alpha}) \). As a comparison, the bound in \( \sigma_0 \) is nontrivial when \( nS(\rho_1\|\rho_0) \leq 2 \). Our result works in a much wider range when \( \alpha \ll 0.5 \), which is desired experimentally. In general, \( M_1 \) is not necessarily a projector, and the explicit form of \( \sigma_0 \) is usually unknown. Nonetheless, it is upper bounded by the the sub-Gaussian norm of a projector under the same \( H_0 \) and with the same false positive rate \( \alpha \) \cite{11}. Intuitively, this results from the fact that for all random variables bounded in \([0, 1]\) and with the same mean, the Bernoulli spreads out the most.

Example.— Let us consider an illustrative example. Suppose the system we are interested in can be in either \( \rho_0 = \frac{1}{2} |0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2} |1\rangle\langle 1| \otimes |1\rangle\langle 1| \) or \( \rho_1 = \frac{m-1}{2m} |0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{m}{2m} |1\rangle\langle 1| \otimes |1\rangle\langle 1| \), where \( |i\rangle \otimes |i\rangle = (|i\rangle)\otimes |i\rangle \) for \( i = 0, 1 \). One can calculate that \( S(\rho_1\|\rho_0) \gtrsim \frac{1}{2m^2} \). When \( m \gg 1 \), the difference between \( \rho_0 \) and \( \rho_1 \) is small. If a projective measurement is performed on \( n \) independent copies of the system, then our result \( \sigma_0 \) states that \( \alpha + \beta \gtrsim 1 - \sqrt{\frac{\alpha - 0.5}{\ln(\alpha/(1-\alpha))} m} \), and if we further control \( \text{Tr}(\rho_0 \otimes n M_1) = \alpha \ll 0.5 \) by some \( M_1 \), then we have \( \beta \gtrsim 1 - \sqrt{n\ln(\alpha/(1-\alpha))} \), which gives us a sense how the type II error rate \( \beta \) depends on the type I error rate \( \alpha \), the number of independent copies of the system \( n \), and the dimension of the system \( m \). If \( \alpha \leq 0.001 \), the system dimension is \( m = 100 \), and we want the lower bound of \( \beta \) no greater than 0.1, then we have to make at least \( n \approx 11.2 \times 10^4 \) copies of the system to achieve this. As a comparison, \( \sigma_0 \) predicts that more than \( n \approx 3.2 \times 10^4 \) copies are needed, a severe underestimate. Hence our result is more informative and provides a more accurate way to estimate the experimental cost.

Discussions.— The generality of our results make them possible to serve as the starting point for various problems including but not limited to quantum hypothesis testing; some of them are discussed below.

Thermodynamic inference.— It turns out that \( \sigma_0 \) can provide insights to thermodynamic inference \cite{15}. By replacing the quantum relative entropy with the Kullback–Leibler divergence, it now reads \( \alpha + \beta \geq 0 - \sigma_0 \sqrt{2nD_{K}(P_1\|P_0)} \). In the classical case, the probabilities can be associated with a forward/backward process in the context of stochastic thermodynamics \cite{15}. Specifically, when \( P_0 \) describes path probability of a backward process in a nonequilibrium steady state, and \( P_1 \) is for the forward process, then \( D_{KL}(P_1\|P_0) = \Delta S \), which is the entropy production in the forward process. The physical meaning of \( \alpha + \beta \geq 1 - \sigma_0 \sqrt{2n\Delta S} \) is that, given \( n \) observed trajectory data, if one wants to infer the arrow of time, then the unavoidable error rate is lower bounded by \( 1 - \sqrt{n\Delta S}/2 \) no matter what \( \Phi \) is used in the test, since \( \sigma_0 \leq 1/2 \) always holds. When the system is almost at equilibrium and \( \Delta S \approx 0 \), then \( \alpha + \beta \approx 1 \), and it is practically impossible to tell the arrow of time. In addition to existing work on estimating the entropy product quantitatively \cite{25, 26}, this result addresses the problem of thermodynamic inference in a more qualitative way.

Speed limit.— Let \( \mathcal{O} \) be an arbitrary observable of interest, \( \mathcal{H} \) be the Hamiltonian, \( \gamma_0 = \rho_1 \) be the density operator of a system at time \( t \), and \( \gamma_1 = \rho_{1+dt} \). Our result \( \sigma_0 \) implies that \( \text{Tr}(\rho_{1+dt}\mathcal{O}) - \text{Tr}(\rho_1\mathcal{O}) \leq \sigma_0 \sqrt{2S(\rho_{1+dt}\|\rho_1)} \), where \( \sigma_0 \) is the sub-Gaussian norm of \( \mathcal{O} \) under \( \rho_1 \). This result does not assume that the dynamics of \( \rho_1 \) is differentiable. If we further assume \( \rho_1 \) is under unitary dynamics, then to second order approximation, \( S(\rho_{1+dt}\|\rho_1) \approx \frac{d^2}{dt^2} \text{Tr}(\rho_1^{-1}C) \) with \( C = -i[\mathcal{H}, \rho_1] \) (Planck’s constant \( \hbar = 1 \)), and \( \sigma_0 \) implies that \( \langle \mathcal{O} \rangle_{t} \leq \sigma_0 \sqrt{\text{Tr}(C \rho_1^{-1}C)} \), where \( \langle \mathcal{O} \rangle_{t} \) denotes the speed of \( \text{Tr}(\rho_1\mathcal{O}) \). This result is different from others such as the Mandelstam-Tamm bound \cite{27}, which states that \( \langle \mathcal{O} \rangle_{t} \leq 2(\Delta_{\mathcal{O}}(\Delta_{\mathcal{H}})) \), where \( \Delta_{\mathcal{O}} \) and \( \Delta_{\mathcal{H}} \) are standard deviations of \( \mathcal{O} \) and \( \mathcal{H} \) with respect to \( \rho_1 \), respectively. For example, when the system is at a state with, say, \( \rho_1 = \frac{1}{2}(|n_1\rangle\langle n_1| + |n_2\rangle\langle n_2|) \) where \( |n_{1,2}\rangle \) are two eigenstates of \( \mathcal{H} \), then \( \langle \mathcal{O} \rangle_{t} = 0 \), and our result shows that \( \langle \mathcal{O} \rangle_{t} \neq 0 \), while both \( \Delta_{\mathcal{O}} \) and \( \Delta_{\mathcal{H}} \) are not necessarily zero, hence our result is better than the Mandelstam-Tamm bound.

There are certainly other speed limit bounds (under different settings) \cite{10}, and it will be an interesting future work to systematically compare our result with them in various physical systems.

Conclusion.— We uncover a general result to bound the mean difference of an arbitrary observable at two arbitrarily different states. This bound involves the quantum relative entropy between two states. When the sub-Gaussian property can be taken advantage of, such a bound can be explicitly expressed by using the sub-Gaussian norm of the observable. We apply this to quantum hypothesis testing and establish a novel bound for the sum of statistical errors, which takes measurement into account and is stronger and more informative than the bound based on quantum Pinsker’s inequality. Applications of our results in some problems of potential interest are discussed. We believe a versatile tool is established in this work for research in various fields.

Acknowledgments.— Y.W. gratefully thank Prof. Dan Netleton and Prof. Huaqing Wu for helpful discussions, and Prof. Thomas Iademola for reviewing an earlier version of this work. Y.W. was supported by the US National Science Foundation under grant HDR: TRIPODS 19-34884.
[1] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer, New York, 1998), 2nd ed.

[2] P. L. Bartlett, O. Bousquet, and S. Mendelson, *Local Rademacher complexities*, Ann. Statist. **33**, 1497 (2005).

[3] R. W. Keener, *Theoretical Statistics: Topics for a Core Course* (Springer, New York, 2010).

[4] Y. Wang, *Sub-Gaussian Error Bounds for Hypothesis Testing* (IEEE International Symposium on Information Theory, Melbourne, 2021).

[5] D. Russo and J. Zou, *How much does your data exploration overfit? Controlling bias via information usage*, IEEE Trans. Inf. Theory, **66**, 302 (2020).

[6] A. Xu and M. Raginsky, *Information-theoretic analysis of generalization capability of learning algorithms* (Advances in Neural Information Processing Systems, Long Beach, 2017).

[7] A. Dechant and S.-i. Sasa, *Fluctuation-response inequality out of equilibrium*, Proc. Natl. Acad. Sci. U.S.A. **117**, 6430 (2020).

[8] M. Hayashi, *Quantum Information Theory: A Mathematical Foundation* (Springer-Verlag, Berlin, 2017), 2nd Ed.

[9] U. Seifert, *From Stochastic Thermodynamics to Thermodynamic Inference*, Annu. Rev. Condens. Matter Phys. **10**, 171 (2019).

[10] L. P. García-Pintos, S. B. Nicholson, J. R. Green, A. del Campo, and A. V. Gorshkov, *Unifying Quantum and Classical Speed Limits on Observables*, Phys. Rev. X **12**, 011038 (2022).

[11] All supporting details and discussions can be found in Supplemental Material.

[12] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge UK, 2000).

[13] C. J. Thompson, *Inequality with Applications in Statistical Mechanics*, J. Math. Phys. **6**, 1812 (1965).

[14] S. Golden, *Lower Bounds for the Helmholtz Function*, Phys. Rev. **137**, B1127 (1965).

[15] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, Rep. Prog. Phys. **75**, 126001 (2012).

[16] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint* (Cambridge University Press, New York, 2019).

[17] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science* (Cambridge University Press, New York, 2018).

[18] V. V. Buldygin and K. K. Moskwichova, *The sub-Gaussian norm of a binary random variable*, Theor. Prob. Math. Statist. **86**, 33 (2013).

[19] Y. Wang, *Sub-Gaussian and subexponential fluctuation-response inequalities*, Phys. Rev. E **102**, 052105 (2020).

[20] K. M. R. Audenaert, J. Calsamiglia, R. Munoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, *Discriminating states: The Quantum Chernoff Bound*, Phys. Rev. Lett. **98**, 160501 (2007).

[21] M. Nussbaum and A. Szkoła, *The Chernoff lower bound for symmetric quantum hypothesis testing*, Ann. Statist. **37**, 1040 (2009).

[22] T. Ogawa and H. Nagaoka, *Strong Converse and Stein’s Lemma in the Quantum Hypothesis Testing*, IEEE Trans. Inf. Theo. **46**, 2428 (2000).

[23] F. Hiai, M. Ohya, and M. Tsukada, *Sufficiency, KMS condition and relative entropy in von Neumann algebras*, Pac. J. Math. **96**, 99 (1981).

[24] J. Arbel, O. Marchal and H. D. Nguyen, *On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables*, ESAIM: Prob. Statist. **24**, 39 (2020).

[25] S. Otsuto, S. Ito, A. Dechant, and T. Sagawa, *Entropy production estimation with optimal current*, Phys. Rev. E **101**, 042138 (2020).

[26] L. Mandelstam and I. Tamm, *The Uncertainty Relation between Energy and Time in Nonrelativistic Quantum Mechanics*, J. Phys. (USSR) **9**, 1 (1945).
Supplemental Material

There are mainly three parts of this file. First, we include the detailed steps to get the quantum fluctuation-response inequality (QFRI) and how to use it in quantum hypothesis testing. Second, we show how to upper bound the sub-Gaussian norm. Third, we provide more details related to thermodynamic inference and speed limit in the Discussion part of the main text.

Derivation of the quantum fluctuation-response inequality

Rather than directly prove the QFRI and its sub-Gaussian version, we start from a more general setting by taking the Bayesian case into account and get back to the main results as special cases. The following two inequalities play an important role in our proof.

**Lemma 1.** *(Klein’s inequality)* Let $A$ and $B$ be Hermitian matrices of the same size, and $f$ be a differentiable convex function defined on the union of the supports of $A$ and $B$. Then we have

\[
\text{Tr}[f(B)] \geq \text{Tr}[f(A)] + \text{Tr}[f'(A)(B - A)].
\]  

**Lemma 2.** *(The Golden-Thompson inequality)* Let $A$ and $B$ be Hermitian matrices, then

\[
\text{Tr}(e^{A+B}) \leq \text{Tr}(e^A e^B).
\]  

Here we will consider the general case that $\pi_0, \pi_1 \in \mathbb{R}$. For an arbitrary Hermitian operator $O$ of typical interest, two density operators $\gamma_0, \gamma_1$, and $s > 0$, let us define

\[
B = \xi s [\pi_1 O - \pi_0 \text{Tr}(\gamma_0 O) I] + \ln \gamma_0,
\]  

\[
A = s \ln \gamma_1 + \text{Tr}(\gamma_1 B - \gamma_1 \ln \gamma_1) I,
\]

where

\[
\xi = \text{sgn}(\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)).
\]

Note $\text{Tr}(\gamma_1 I) = \text{Tr}(\gamma_1) = 1$, then it is straightforward to check that

\[
\text{Tr}(\gamma_1 (B - A)) = \text{Tr}(\gamma_1 B) - \text{Tr}(\gamma_1 \ln \gamma_1) - \text{Tr}(\gamma_1 B - \gamma_1 \ln \gamma_1) \text{Tr}(\gamma_1 I) = 0.
\]

We will denote $\lambda = \text{Tr}(\gamma_1 B - \gamma_1 \ln \gamma_1)$, so $A = s \ln \gamma_1 + \lambda I$.

First, by Klein’s inequality where $f(x) = e^x$ is selected, then $f'(x) = e^x$, and we have

\[
\text{Tr}(e^B) \geq \text{Tr}(e^A) + \text{Tr}(e^A(B - A))
\]

\[
= \text{Tr}[\exp(\ln \gamma_1 + \lambda I)] + \text{Tr}[\exp(\ln \gamma_1 + \lambda I)(B - A)]
\]

\[
= \text{Tr}(\gamma_1 e^{\lambda I}) + \text{Tr}(\gamma_1 e^{\lambda I}(B - A))
\]

\[
= \text{Tr}(\gamma_1 e^{\lambda I}) + \text{Tr}(\gamma_1 e^{\lambda I} B)
\]

\[
= e^{\lambda} \text{Tr}(\gamma_1) + e^{\lambda} \text{Tr}(\gamma_1 B)
\]

\[
= e^{\lambda},
\]

by which we immediately obtain

\[
\ln \text{Tr}(e^B) \geq \lambda = \text{Tr}(\gamma_1 B - \gamma_1 \ln \gamma_1)
\]

\[
= \text{Tr}(\gamma_1 \xi s [\pi_1 O - \pi_0 \text{Tr}(\gamma_0 O) I] + \ln \gamma_0) - \text{Tr}(\gamma_1 \ln \gamma_1)
\]

\[
= \xi s [\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O) \text{Tr}(\gamma_1 I)] + \text{Tr}(\gamma_1 \ln \gamma_0) - \text{Tr}(\gamma_1 \ln \gamma_1)
\]

\[
= \xi s [\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)] - S(\gamma_1 \| \gamma_0).
\]

\[
= s |\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)| - S(\gamma_1 \| \gamma_0).
\]

\[
= s |\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)| - S(\gamma_1 \| \gamma_0).
\]  

[by the definition of $S(\gamma_1 \| \gamma_0)$ and $\text{Tr}(\gamma_1) = 1$]  

[by the definition of $\xi$]  

(14)
On the other hand, by (11), we note

\[
\ln \text{Tr}(e^B) = \ln \text{Tr} [\exp(\xi s [\pi_1 O - \pi_0 \text{Tr}(\gamma_0 O) I] + \ln \gamma_0)] \\
\leq \ln \text{Tr} [\gamma_0 \exp(\xi s [\pi_1 O - \pi_0 \text{Tr}(\gamma_0 O) I])] \\
= \ln \text{Tr} [\gamma_0 \exp(\xi s [\pi_1 O - \pi_0 \text{Tr}(\gamma_0 O) I] + \pi_1 \text{Tr}(\gamma_0 O) I)] \\
= \ln \text{Tr} [\gamma_0 e^{\xi \sqrt{\pi_1} O_c + (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) I}] \\
= \ln \text{Tr} [\gamma_0 e^{\xi \sqrt{\pi_1} O_c} e^{\xi s(\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) I}] \\
= \ln \text{Tr} [\gamma_0 e^{\xi \sqrt{\pi_1} O_c} e^{\xi s(\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) I}] \\
= \ln \{ e^{\xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) I} \} \\
= \ln \text{Tr} (\gamma_0 e^{\xi s \pi_1 O_c}) + \xi s(\pi_1 - \pi_0) \text{Tr}(\gamma_0 O).
\]

(15)

Combining (14) and (15), and noting \( s > 0 \), we have

\[
|\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)| \leq \frac{1}{s} \ln \text{Tr} (\gamma_0 e^{\xi s \pi_1 O_c}) + \xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) + \frac{1}{s} S(\gamma_1 || \gamma_0).
\]

(16)

Since (16) holds for any \( s > 0 \) where each term is well defined, and for typical operators of interest in quantum information science it holds for all \( s > 0 \), hence we have

\[
|\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)| \leq \inf_{s > 0} \left[ \frac{1}{s} \ln \text{Tr} (\gamma_0 e^{\xi s \pi_1 O_c}) + \xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) + \frac{1}{s} S(\gamma_1 || \gamma_0) \right].
\]

(17)

Using an orthonormal basis \( \{|i\} \) that diagonalizes \( O_c \), we further have

\[
\text{Tr} (\gamma_0 e^{\xi s \pi_1 O_c}) = \text{Tr} \left( \gamma_0 \sum_i |i\rangle \langle i| e^{\xi s \pi_1 O_c} \right) = \sum_i \langle i| \gamma_0 |i\rangle e^{\xi s \pi_1 O_c} \equiv \sum_i p_i e^{\xi s \pi_1 O_c},
\]

where \( \{o_i\} \) are eigenvalues of \( O_c \). This defines a random variable \( O \) that equals \( o_i \) with probability \( p_i \). \( O \) has mean zero under \( P \equiv \{p_i\} \) since

\[
\mathbb{E}_{O \sim P} O = \text{Tr}(\gamma_0 O_c) = \text{Tr}(\gamma_0 O) - \text{Tr}(\gamma_0 O) \text{Tr}(I) = 0.
\]

If \( O \) is bounded, then \( O_c \) is sub-Gaussian in the following sense: by (5), we have

\[
\text{Tr} (\gamma_0 e^{\xi s \pi_1 O_c}) = \mathbb{E}_{O \sim P} e^{\xi s \pi_1 O} \leq e^{\sigma_{OP}^2 \xi^2 / 2} = e^{(\pi_1 \sigma_{OP})^2 s^2 / 2},
\]

(18)

where \( \sigma_{OP} \) is the sub-Gaussian norm of \( O \) under \( P \). And (17) can be further written as

\[
|\pi_1 \text{Tr}(\gamma_1 O) - \pi_0 \text{Tr}(\gamma_0 O)| \leq \inf_{s > 0} \left[ \frac{1}{s} \ln \text{Tr} (\gamma_0 e^{\xi s \pi_1 O_c}) + \xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) + \frac{1}{s} S(\gamma_1 || \gamma_0) \right] \\
\leq \inf_{s > 0} \left[ \frac{1}{2s} (\pi_1 \sigma_{OP})^2 s^2 + \xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) + \frac{1}{s} S(\gamma_1 || \gamma_0) \right] \\
= \inf_{s > 0} \left[ \frac{1}{2} (\pi_1 \sigma_{OP})^2 s + \xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O) + \frac{1}{s} S(\gamma_1 || \gamma_0) \right] \\
= \sigma_{OP} \sqrt{2S(\gamma_1 || \gamma_0)} + \xi (\pi_1 - \pi_0) \text{Tr}(\gamma_0 O),
\]

(19)

where the infimum is achieved at \( s = \sqrt{2S(\gamma_1 || \gamma_0) / \pi_1 \sigma_{OP}} \).

Now we take a look at two special cases of (19).

- \( \pi_0 = \pi_1 = 1 \), then the sub-Gaussian quantum fluctuation-response inequality (5) is recovered.

- Let \( \pi_0, \pi_1 \in (0, 1) \) and \( O = M_1 \) as an element of a POVM. Note type I error rate \( \alpha = \text{Tr}(\gamma_0 M_1) \) and type II error rate \( \beta = \text{Tr}(\gamma_1 M_0) = 1 - \text{Tr}(\gamma_1 M_1) \), then we have

\[
|\pi_1 (1 - \beta) - \pi_0 \alpha| \leq \pi_1 \sigma_{OP} \sqrt{2S(\gamma_1 || \gamma_0)} + \xi (\pi_1 - \pi_0) \alpha,
\]

\[
\Rightarrow \pi_1 \left[ 1 - \sigma_{OP} \sqrt{2S(\gamma_1 || \gamma_0)} \right] - \xi (\pi_1 - \pi_0) \alpha \leq \pi_0 \alpha + \pi_1 \beta \leq \pi_1 \left[ 1 + \sigma_{OP} \sqrt{2S(\gamma_1 || \gamma_0)} \right] + \xi (\pi_1 - \pi_0) \alpha.
\]

(20)
When $\pi_0 = \pi_1 = 1/2$ and let $\sigma_0 \equiv \sigma_{\text{OP}}$, then our error bound is recovered. In general, however, this result depends on $\xi$. In the case that $\xi = 1$, we have
$$
\pi_1 \text{Tr}(\gamma_1 \mathcal{M}_1) > \pi_0 \text{Tr}(\gamma_0 \mathcal{M}_1) \implies \pi_0 \alpha + \pi_1 \beta < \pi_1.
$$

Under such a constraint, we obtain lower and upper bounds that
$$
\alpha + \beta \geq 1 - \sigma_{\text{OP}} \sqrt{2S(\gamma_1 \| \gamma_0)} \quad \text{and} \quad \left(\frac{2\pi_0}{\pi_1} - 1\right) \alpha + \beta \leq 1 + \sigma_{\text{OP}} \sqrt{2S(\gamma_1 \| \gamma_0)}.
$$

Similarly, for $\xi = -1$, the constraint is $\pi_0 \alpha + \pi_1 \beta > \pi_1$, and we have error bounds
$$
\left(\frac{2\pi_0}{\pi_1} - 1\right) \alpha + \beta \geq 1 - \sigma_{\text{OP}} \sqrt{2S(\gamma_1 \| \gamma_0)} \quad \text{and} \quad \alpha + \beta \leq 1 + \sigma_{\text{OP}} \sqrt{2S(\gamma_1 \| \gamma_0)}.
$$

These bounds provide new insights to the sum of error rates in the Bayesian setting.

**Upper bounding the sub-Gaussian norm**

We will show that the sub-Gaussian norm of $\mathcal{M}_1$ under $H_0 : \rho \sim \gamma_0$ as defined in the main text can be upper bounded by that of a $\{0, 1\}$-valued test function (a Bernoulli random variable) with the same $\alpha$. Recall that we use a set of orthonormal base vectors $\{|m\rangle\}$ to diagonalize $\mathcal{M}_1$, and
$$
\alpha = \text{Tr}(\gamma_0 \mathcal{M}_1) = \text{Tr} \left( \gamma_0 \sum_m |m\rangle \langle m| \mathcal{M}_1 \right) = \sum_m p_m \mu_m = \sum_{m: \mu_m > 0} p_m \mu_m,
$$

where $p_m = \langle m | \gamma_0 | m \rangle$ and $\mu_m \in [0, 1]$ is the eigenvalue of $\mathcal{M}_1$ corresponding to $|m\rangle$. This procedure induces a random variable $M$ that takes on values in $\{\mu_m\}$ with corresponding the probability $P_M = \{p_m\}$. Using $M$, we have
$$
\text{Tr}(\gamma_0 e^{\xi M} \mathcal{M}_1) = \mathbb{E}_{M \sim P_M} e^{\xi M} = \sum_m p_m e^{\xi \mu_m} = \left( \sum_{m: \mu_m = 0} + \sum_{m: \mu_m > 0} \right) p_m e^{\xi \mu_m} = 1 - \sum_{m: \mu_m > 0} p_m + \sum_{m: \mu_m > 0} p_m e^{\xi \mu_m} = 1 + \sum_{m: \mu_m > 0} p_m \left(e^{\xi \mu_m} - 1\right).
$$

On the other hand, let us define a Bernoulli random variable $Y \sim P_Y = \text{Bernoulli}(\alpha)$. That is, $Y = 0$ with probability $1 - \alpha$ and $Y = 1$ with probability $\alpha$. Note $Y$ and $M$ have the same mean $\alpha$, and
$$
\mathbb{E}_{Y \sim P_Y} e^{\xi Y} = (1 - \alpha) + \alpha e^{\xi} = 1 + \alpha \left(e^{\xi} - 1\right) = 1 + \sum_{m: \mu_m > 0} p_m \left(\mu_m e^{\xi} - \mu_m\right).
$$

The difference between $\mathbb{E}_{Y \sim P_Y} e^{\xi Y}$ and $\mathbb{E}_{M \sim P_M} e^{\xi M}$ can then be calculated as
$$
\mathbb{E}_{Y \sim P_Y} e^{\xi Y} - \mathbb{E}_{M \sim P_M} e^{\xi M} = \sum_{m: \mu_m > 0} p_m \left(\mu_m e^{\xi} - \mu_m - e^{\xi \mu_m} + 1\right) \equiv \sum_{m: \mu_m > 0} p_m f(\mu_m, s),
$$

where $f(\mu, s)$ is a function that captures the difference in the expectations.
where \( f(x, s) \equiv xe^{\xi s} - e^{\xi sx} + (1 - x) \) for \( x \in [0, 1] \) and \( s > 0 \). Note that
\[
\frac{\partial f(x, s)}{\partial s} = \xi xe^{\xi s} - \xi xe^{\xi sx}.
\]
If \( \xi = 1 \), then
\[
\frac{\partial f(x, s)}{\partial s} = xe^s - xe^{sx} = x(e^s - e^{sx}) \geq 0
\]
since \( x \in [0, 1] \) and thus \( e^s \geq e^{sx} \) for \( s > 0 \). If \( \xi = -1 \), then
\[
\frac{\partial f(x, s)}{\partial s} = -xe^{-s} + xe^{-sx} = x(e^{-sx} - e^{-s}) \geq 0
\]
since \( e^{-sx} \geq e^{-s} \) when \( x \in [0, 1] \) and \( s > 0 \). Hence we conclude that
\[
\frac{\partial f(x, s)}{\partial s} \geq 0,
\]
which indicates that \( f \) is increasing in \( s \). Also note \( f(x, 0) = 0 \). We thus finally arrive at
\[
f(x, s) \geq 0,
\]
which implies that
\[
\mathbb{E}_{M \sim P_M} e^{\xi s(M - \alpha)} \leq \mathbb{E}_{Y \sim P_Y} e^{\xi s(Y - \alpha)} \leq e^{\sigma_Y^2 s^2/2},
\]
where \( \sigma_Y \) is the sub-Gaussian norm of the binary random variable \( Y \), which is known to be \( \sigma_Y = \sqrt{(0.5 - \alpha)/\ln((1 - \alpha)/\alpha)} \).
Since by definition, the sub-Gaussian norm of \( M \) is an infimum, hence \( \sigma_M \leq \sigma_Y \). We thus have proved that the sub-Gaussian norm of \( M_1 \) under \( H_0 \) is upper bounded by \( \sqrt{(0.5 - \alpha)/\ln((1 - \alpha)/\alpha)} \).

**Discussions: thermodynamic inference and speed limit**

First, let us consider thermodynamic inference in the setting of stochastic thermodynamics. At a nonequilibrium steady state, let \( \omega \) be a stochastic trajectory of a Brownian particle in a time interval. Let \( \omega^\dagger \) be the time-reversed trajectory. One can establish a one-to-one map \( I \) between these two, say, \( \omega^\dagger = I(\omega) \) for any \( \omega \). Suppose \( P(\omega) \) is the probability measure for \( \omega \), and \( P^\dagger(\omega^\dagger) \) is that for \( \omega^\dagger \). Then \( I \) induces the pushforward measure \( I\#P \) such that
\[
I\#P(\omega) = P^\dagger(\omega^\dagger).
\]
By standard arguments in stochastic thermodynamics, the mean entropy production in the forward process \( \Delta S \) can thus be represented by
\[
\Delta S = D_{KL}(P(\omega)||P^\dagger(\omega^\dagger)) = D_{KL}(P(\omega)||I\#P(\omega)).
\]
Given \( n \) observations of trajectory data \( \{\omega_i\}_{i=1}^n \), one aims to compare the empirical distribution of \( \omega_i \) (or equivalently, of some summary statistic of \( \omega_i \)) with \( P \) and \( I\#P \), and this falls into the framework of classical hypothesis testing. By defining a test function \( \Phi \), which is the function of the data, \( P \), and \( I\#P \), and the range of which is \( \{0, 1\} \), one inevitably has the two types of statistical errors \( \alpha \) and \( \beta \). Thus a direct correspondence to the quantum error bound in this case is
\[
\alpha + \beta \geq 1 - \sigma_0 \sqrt{2n\Delta S},
\]
where \( \sigma_0 \) is the sub-Gaussian norm of \( \Phi \) under \( H_0 : I\#P \) generates the data. (Oftentimes, \( H_0 \) is the one that we want to reject.) Since \( \Phi \) is a Bernoulli random variable, and hence sub-Gaussian, \( \sigma_0 \) is always less than 0.5 unless \( \alpha = 0.5 \). So we can conclude that \( \alpha + \beta \geq 1 - \sqrt{n\Delta S/2} \) even without knowing the details of \( \Phi \). This result reflects the intrinsic difficulty in identifying time’s arrow near equilibrium where \( \Delta S \approx 0 \); any test function used in the hypothesis testing is no better than random guess.
Then, let us consider the speed limit example, where $\rho_t$ is under unitary dynamics. That is, $\rho_{t+dt} = e^{-iHdt}\rho_te^{iHdt}$. Note that, to second order, we have

$$\rho_{t+dt} \approx \left(1 - iHdt - \frac{1}{2}H^2dt^2\right) \rho_t \left(1 + iHdt - \frac{1}{2}H^2dt^2\right)$$

$$= \rho_t - i[H, \rho_t]dt - \frac{1}{2}H^2\rho_t dt^2 + i\rho_t H dt - \frac{1}{2}\rho_t H^2 dt^2 + H\rho_t H dt^2$$

$$= \rho_t \left(I - i\rho_t^{-1}[H, \rho_t]dt + \frac{1}{2}(2\rho_t^{-1}H\rho_t H - \rho_t^{-1}H^2\rho_t - \rho_t^{-1}\rho_t H^2)dt^2\right)$$

$$= \rho_t \left(I - i\rho_t^{-1}[H, \rho_t]dt + \frac{1}{2}(2\rho_t^{-1}H\rho_t H - \rho_t^{-1}H^2\rho_t - \rho_t^{-1}\rho_t H^2)dt^2\right)$$

where $\rho_t$ and $\hat{I}$ are positive definite. Hence by the property of positive definite operators, we have

$$\text{Tr}(\rho_{t+dt} \ln \rho_{t+dt}) \approx \text{Tr}(\rho_{t+dt} \ln(\rho_t \hat{I})) = \text{Tr}(\rho_{t+dt} \ln \rho_t) + \text{Tr}(\rho_{t+dt} \ln \hat{I}).$$

As a result, we can approximate $S(\rho_{t+dt} || \rho_t)$ to second order as

$$S(\rho_{t+dt} || \rho_t) = \text{Tr}(\rho_{t+dt} \ln \rho_{t+dt}) - \text{Tr}(\rho_{t+dt} \ln \rho_t)$$

$$\approx \text{Tr}(\rho_{t+dt} \ln \hat{I})$$

$$\approx \text{Tr}(\rho_{t+dt}(-i[H, \rho_t]dt + \frac{1}{2}(2\rho_t^{-1}H\rho_t H - \rho_t^{-1}H^2\rho_t - \rho_t^{-1}\rho_t H^2)dt^2))$$

$$\approx \text{Tr}((-i[H, \rho_t])dt + \frac{1}{2}\rho_t H^2 dt^2 - \frac{1}{2}\rho_t H^2 dt^2 - \frac{1}{2}[H, \rho_t]\rho_t^{-1}[H, \rho_t]dt^2)$$

$$= -i\text{Tr}((-i[H, \rho_t])dt + \frac{1}{2}\text{Tr}((-i[H, \rho_t])\rho_t^{-1}(-i[H, \rho_t]) dt^2)$$

$$= \frac{1}{2}\text{Tr}(C\rho_t^{-1}) dt^2.$$ 

Note $C \equiv (-i[H, \rho_t]) = (-i)(H\rho_t - \rho_t H) = (-i)(\rho_t H - \rho_t H) = (i[\rho, H]) = (i[\rho, H])^\dagger = (i[\rho, H]) = C^\dagger$. Hence $C$ is Hermitian, and $\text{Tr}(C\rho_t^{-1}) = \text{Tr}(C\rho_t^{-1/2})(C\rho_t^{-1/2})^\dagger \geq 0$. This leads to our conclusion that $|\langle \hat{O} \rangle| \leq \sigma_{\text{OT}}\sqrt{\text{Tr}(C\rho_t^{-1})}.$

As a simple example to show the difference between our bound and the Mandelstam-Tamm bound, let us consider a two-level system with $\mathcal{H} = n_1|n_1\rangle\langle n_1| + n_2|n_2\rangle\langle n_2|$ and $\rho_t = \frac{1}{2}|n_1\rangle\langle n_1| + \frac{1}{2}|n_2\rangle\langle n_2|$. It is straightforward to check that $[\rho_t, \mathcal{H}] = 0$, hence $C = 0$, and our bound suggests $\langle \hat{O} \rangle_t = 0$. On the other hand, $\text{Tr}(\mathcal{H}^2\rho_t) = \frac{1}{2}(n_1^2 + n_2^2)$ and $\text{Tr}(\mathcal{H}\rho_t) = \frac{1}{4}(n_1 + n_2)$, and thus $\Delta_t \mathcal{H} = \frac{1}{2}|n_1 - n_2|$. In general $\Delta_t \mathcal{O} > 0$, hence the Mandelstam-Tamm bound is a positive number, and not as tight as ours.