Sub-Gaussian and sub-exponential fluctuation-response inequalities

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

Sub-Gaussian and sub-exponential distributions are introduced and applied to study the fluctuation-response relation out of equilibrium. A bound on the difference in expected values of an arbitrary sub-Gaussian or sub-exponential physical quantity is established in terms of its sub-Gaussian or sub-exponential norm. Based on that, we find the entropy difference between two states is bounded by the energy fluctuation in these states. Moreover, we obtain generalized versions of the thermodynamic uncertainty relation in different regimes. Some operational issues are also addressed, non-asymptotic bounds on the errors incurred by using the sample mean instead of the expected value in our fluctuation-response inequalities are derived.

I. INTRODUCTION

Gaussian distributions play a crucial role in statistical physics. As a classical example, when a thermodynamical system is at equilibrium, a typical physical quantity fluctuates around its ensemble average in a Gaussian way [11]. Two amazing properties of a Gaussian random variable are that (i) it can be fully characterized by its mean and variance, without resort to higher order statistics; (ii) the Gaussian property is preserved under linear transformation. Thanks to such properties, the highly successful linear response theory can be established [2,3], which relies heavily on the assumption that when the perturbation is weak, the deviation from the original equilibrium state is small, and one can work in the regime where the leading effect of an external force is linear. However, when the perturbation is strong and the nonlinear effect has to be taken into account, the linear response theory is no long working well. Recent years have witnessed substantial advances in nonequilibrium statistical physics. Results such as Jarzynski’s equality [4] and various kinds of fluctuation relations [5] have been shown to be valid for quite general nonequilibrium processes, beyond the linear response regime. Such theories consider the system’s evolution path in phase space, and associate each trajectory with some physical quantities like (stochastic) work and entropy production that are defined in an unusual sense. In an experiment that manipulates a single molecule of RNA between two conformations, it is found that when the perturbation is weak, the distribution of trajectory-dependent dissipated work can be well approximated by a Gaussian distribution, but this is not true for stronger perturbations [6]. Actually, a large body of literature exists regarding the non-Gaussian distributions encountered in nonequilibrium physics, both theoretically and experimentally; for example, see Refs. [7–12].

Despite being non-Gaussian, it seems that these distributions typically are unimodal, and they differ from a Gaussian mainly in the existence of skewness and a different decay rate in the tail probability. One might correct such deviations by considering higher order statistics [8–14], but that may require the dynamics of the system, or the “perturbation” may be too strong to be treated in a perturbative way. In this work, we introduce the sub-Gaussian and sub-exponential distributions as two classes of distributions that are particularly relevant to nonequilibrium physics. They play an important role in modern statistics and machine learning [15,16], but seem to be less known to the physics community. Roughly speaking, sub-Gaussian distributions are those possessing a tail that uniformly falls under a curve which is the tail of some Gaussian distribution lifted up, and sub-exponential distributions are those possessing a tail dominated by the uplifted tail of some exponential distribution. Gaussian distributions belong to the sub-Gaussian class, and sub-Gaussian distributions belong to the sub-exponential class. Such a hierarchical structure is depicted in Fig. 1. We argue that beyond the linear response regime that corresponds to Gaussian distributions, in the nonlinear regime, sub-Gaussian and sub-exponential distributions are ubiquitous, as previous study shows. As an example, the fluctuation relations have the form \( P(x)/P(-x) = e^x \), where \( x \) represents some measure of irreversibility, its precise meaning depending on context [5]. Note that for all \( x \geq 0 \), we have \( P(X \leq -x) \leq e^{-x} \) [7]. Hence the random variable \( X \) has a tail that decays at least in an exponential way, thus \( X \) is at least one-sided sub-exponential. Another physically relevant fact is that all bounded distributions are sub-Gaussian, hence one expects that sub-Gaussian distributions may be suitable for physical systems with finite states.

Rather than to analyze in detail the dynamics of a specific system, our aim in this work is to study how one can take advantage of the general properties of sub-Gaussian and sub-exponential distributions to study the nonlinear response theory in a unified way for different systems. However, the price we pay for this universality is that often times we can only obtain results in the form of inequalities rather than equalities. This is reminiscent of the fact that the second law of thermodynamics is universally true for macroscopic systems, but the lower bound it provides on the entropy increase in a thermodynamical process can be substantially improved when more detailed information of the system in question is gained. The trade-off between universality and tightness of a bound is inevitable for thermodynamical theories involving inequalities.

To study the fluctuation-response relation out of equilibrium, we mainly follow the idea proposed in a recent work by Dechant and Sasa [18]. By using the so-called sub-Gaussian (or sub-exponential) norm, we are able to further refine their results and provide a neat upper bound for the difference between expected values of an arbitrary sub-Gaussian (or sub-
properties of sub-Gaussian and sub-exponential distributions in numerical analysis. We argue sub-Gaussian and sub-exponential random variables are relevant in general nonequilibrium mechanistical states or processes where nonlinear response theory applies.

In the following, we introduce the concepts and basic properties of sub-Gaussian and sub-exponential random variables, such as the Orlicz $\psi_2$-norm. These norms are strongly related and equivalent to each other up to a numerical constant factor; they emphasize on different aspects of the sub-Gaussian property. The reason why we choose $\|X\|_G$ defined above as the sub-Gaussian norm is that $\|X\|_G^2$ naturally reduces to the variance $\text{var}(X)$ if $X$ is Gaussian. While, in general cases, we have $\|X\|_G^2 \geq \text{var}(X)$. Non-trivial universal results could be obtained for qualitatively similar physical processes, in terms of the corresponding sub-Gaussian norm.

It is worth noting that when we speak of a sub-Gaussian distribution, it does not necessarily mean we refer to a family of distributions with a fixed parametric form like Gaussian distributions, which are parameterized by mean and variance. We can work with a sub-Gaussian variable as long as the condition (1) holds, even the explicit form of the distribution is unknown or intractable. The sub-Gaussian class of distributions is probably the simplest generalization of Gaussian distributions that can be relevant in nonequilibrium statistical physics. Typical distributions that are sub-Gaussian include Gaussian, Bernoulli, and all bounded distributions.

A centered random variable $X$ is sub-Gaussian if for some $\sigma > 0$, its moment generating function $\mathbb{E}e^{sX}$ satisfies

$$\mathbb{E}e^{sX} \leq e^{s^2\sigma^2/2}, \forall s \in \mathbb{R}. \quad (1)$$

Apparently, a Gaussian random variable is also sub-Gaussian. For our purposes, we also define the sub-Gaussian norm of $X$ as the infimum of $\sigma$ such that the sub-Gaussian property (1) holds:

$$\|X\|_G = \inf\{\sigma > 0 : \mathbb{E}e^{sX} \leq e^{s^2\sigma^2/2}, \forall s \in \mathbb{R}\}. \quad (2)$$

$\|X\|_G^2$ is sometimes referred to as the optimal proxy variance. There are other ways to define a norm for sub-Gaussian variables, such as the Orlicz $\psi_2$-norm. These norms are strongly related and equivalent to each other up to a numerical constant factor; they emphasize on different aspects of the sub-Gaussian property. The reason why we choose $\|X\|_G^2$ defined above as the sub-Gaussian norm is that $\|X\|_G^2$ naturally reduces to the variance $\text{var}(X)$ if $X$ is Gaussian. While, in general cases, we have $\|X\|_G^2 \geq \text{var}(X)$. Non-trivial universal results could be obtained for qualitatively similar physical processes, in terms of the corresponding sub-Gaussian norm.

It is worth noting that when we speak of a sub-Gaussian distribution, it does not necessarily mean we refer to a family of distributions with a fixed parametric form like Gaussian distributions, which are parameterized by mean and variance. We can work with a sub-Gaussian variable as long as the condition (1) holds, even the explicit form of the distribution is unknown or intractable. The sub-Gaussian class of distributions is probably the simplest generalization of Gaussian distributions that can be relevant in nonequilibrium statistical physics. Typical distributions that are sub-Gaussian include Gaussian, Bernoulli, and all bounded distributions.

For a centered sub-Gaussian variable $X$, one can establish the concentration inequality, which is used to bound its tail probability. To this end, note $P(X \geq t)$ for all $t \geq 0$ and $s > 0$ can be bounded as

$$P(X \geq t) = P(e^{sX} \geq e^{st}) \leq e^{-st}\mathbb{E}e^{sX},$$

where we have used the Markov’s inequality in the last step. Then the Chernoff bound states that

$$\log P(X \geq t) \leq \inf_{s>0} \{\log \mathbb{E}e^{sX} - st\},$$

FIG. 1. (Color online) A hierarchical structure of distributions. Whereas Gaussian distributions lay the foundation for linear response theory, we argue sub-Gaussian and sub-exponential distributions are relevant in general nonequilibrium mechanistical states or processes where nonlinear response theory applies.
combining which with (1) and (2) we obtain the infimun is achieved at \( s = t/\|X\|_G^2 \), and \( P(X \geq t) \leq e^{-t^2/2\|X\|_G^2} \). Similarly, one can bound \( P(X \leq -t) \), hence we come to the concentration inequality for the sub-Gaussian variable

\[
P(\|X\| \geq t) \leq 2 \exp\left( -\frac{t^2}{2\|X\|_G^2} \right),
\]

which holds for all \( t \geq 0 \).

The sub-Gaussian property is preserved under addition. If \( X_1, \ldots, X_N \) are independent and sub-Gaussian variables with norms \( \|X_1\|_G, \ldots, \|X_N\|_G \), respectively. Let \( X = \sum_{i=1}^N a_i X_i \), where \( a_i \)'s are constants. Then \( X \) is also sub-Gaussian, and we can compute its norm by (1) and (2):

\[
\mathbb{E}e^{sX} = \mathbb{E}e^{s\sum_{i=1}^N a_i X_i} = \prod_{i=1}^N e^{sa_i X_i} \\
\leq \prod_{i=1}^N e^{s^2 a_i^2 \|X_i\|_G^2/2} = e^{s^2 \sum_{i=1}^N a_i^2 \|X_i\|_G^2/2}.
\]

Thus \( X \) is sub-Gaussian with \( \|X\|_G^2 = \sum_{i=1}^N a_i^2 \|X_i\|_G^2 \). By this result, we have the Hoeffding’s inequality that

\[
P\left( \left| \frac{1}{N} \sum_{i=1}^N X_i \right| \geq t \right) \leq 2 \exp\left( -\frac{N^2t^2}{2\sum_{i=1}^N \|X_i\|_G^2} \right),
\]

which holds for all \( t \geq 0 \). Furthermore, if these \( X_i \)'s are identically distributed, then the sample mean \( \bar{X} = \sum_{i=1}^N X_i/N \) satisfies \( P(|\bar{X}| \geq t) \leq 2e^{-Nt^2/2\|\bar{X}\|_G^2} \).

### B. Sub-exponential random variables

Similarly, we briefly provide classical results for centered sub-exponential variables here. Although there is no consensus on, and exist different versions of the definition of sub-exponential variables, these definitions are consistent with each other, all leading to the same kind of probability inequalities. In this work, a centered sub-exponential variable is a zero-mean random variable that satisfies

\[
\mathbb{E}e^{sX} \leq e^{\sigma^2 s^2/2}, \quad \text{for } |s| \leq \frac{c_E}{\sigma},
\]

where \( \sigma > 0 \) and \( c_E = (\sqrt{3} + 1)/2 \) is picked for later convenience. Our definition of sub-exponential variables is chosen in a way that not only ensures \( \|X\|_E = \|X\|_G = \text{var}(X) \) when \( X \) is centered Gaussian, but also assumes a neatest form \( e^{-t/\|X\|_E} \) for the exponentially-decaying tail without additional coefficients, as shown below. Note that different than in (1), here the range of \( s \) is confined, thus apparently, if \( X \) is sub-Gaussian then it is automatically sub-exponential. One non-trivial example of a centered sub-exponential variable is \( X = Z^2 - 1 \), where \( Z \sim N(0, 1) \). Note \( Z^2 \) follows the chi-square distribution with 1 degree of freedom, whose tail decays essentially in an exponential way, and \( \mathbb{E}e^{sX} = e^{Ee^{s(2Z^2-1)}} = e^{s^2e^{-s}-s} = e^{-s}/\sqrt{1-2s} \). Note \( \mathbb{E}e^{sX} \) is only well-defined for \( s < 1/2 \), and numerically one can check that \( \mathbb{E}e^{sX} \leq e^{s^2 2}/2 \) for \( |s| \leq c_E/3 \). Hence \( X \) is sub-exponential.

Next, we define the sub-exponential norm as

\[
\|X\|_E = \inf\{\sigma > 0 : \mathbb{E}e^{sX} \leq e^{\sigma^2 s^2/2}, \text{for } |s| \leq c_E/\sigma\},
\]

Also, there are other possible norms for sub-exponential variables such as the Orlicz \( \psi_1 \)-norm, which are all strongly related and equivalent to each other up to a numerical constant factor.

One can establish the concentration inequality for sub-exponential variables in terms of \( \|X\|_E \). As above, let us start with the Chernoff bound. For \( t \geq 0 \) and \( 0 < s \leq c_E/\|X\|_E \), we have:

\[
\log P(X \geq t) \leq \inf_{0 < s \leq c_E/\|X\|_E} \{\mathbb{E}e^{sX} e^{-st}\}.
\]

Noting (4), (5), and the range of \( s \), we can see that if \( t > c_E/\|X\|_E \) then \( \min_{s} \mathbb{E}e^{\|X\|_E^2 s^2/2 - st} = e^{t^2/2 - c_Et}/\|X\|_E \), which is achieved at the boundary \( s = c_E/\|X\|_E \), and \( P(X \geq t) \leq e^{t^2/2 - c_Et}/\|X\|_E \leq e^{t^2/2 - c_Et}/\|X\|_E \leq e^{-t^2/2}/\|X\|_E \). Apply similar arguments to \( -X \), then we have for \( t > c_E/\|X\|_E \) that \( P(|X| \geq t) \leq 2e^{-t^2/2}/\|X\|_E \). While if \( 0 \leq t \leq c_E/\|X\|_E \), then \( \min_{s} \mathbb{E}e^{\|X\|_E^2 s^2/2 - st} = e^{-t^2/2}/\|X\|_E^2 \), which is achieved at \( s = t/\|X\|_E \). Hence for \( t \) in this range, the situation is the same as in the sub-Gaussian case, and we have \( P(|X| \geq t) \leq 2e^{-t^2/2}/\|X\|_E^2 \). Roughly speaking, one can see that for small \( t \) deviation from the mean 0, the sub-exponential variable actually has no difference from a sub-Gaussian variable. The difference manifests itself for large \( t \) deviation. Combining the results together, we have the concentration inequality that for all \( t \geq 0 \), a centered sub-exponential variable satisfies

\[
P(|X| \geq t) \leq \begin{cases} 
2 \exp\left( -\frac{t^2}{2\|X\|_E^2} \right), & 0 \leq |t| \leq c_E/\|X\|_E; \\
2 \exp\left( -\frac{t^2}{\|X\|_E} \right), & |t| > c_E/\|X\|_E.
\end{cases}
\]

or in a more compact form,

\[
P(|X| \geq t) \leq 2 \exp\left( -\min\left\{ \frac{t^2}{2\|X\|_E^2}, \frac{t^2}{\|X\|_E} \right\} \right).
\]

Slightly different than in the sub-Gaussian case, for \( N \) independent centered sub-exponential variables \( X_1, \ldots, X_N \) with norms \( \|X_1\|_E, \ldots, \|X_N\|_E \), there are two relevant norms associated with \( X = \sum_{i=1}^N a_i X_i \). As before, for \( t \geq 0 \) and \( 0 < s \leq 1/\max_{i}\{a_i \|X_i\|_E\} \), we have

\[
P(X \geq t) \leq e^{-st} \mathbb{E}e^{\sum_{i=1}^N a_i X_i} \leq e^{-st} \prod_{i=1}^N e^{s^2a_i^2 \|X_i\|_E^2/2} = e^{s^2 \sum_{i=1}^N a_i^2 \|X_i\|_E^2/2 - st}.
\]

Hence, we can perform a similar analysis to the single variable case. First, if each \( X_i \) is in the sub-Gaussian regime for small \( t \), then \( X \) is sub-Gaussian-like with squared norm \( \|X\|_E^2 = \sum_{i=1}^N a_i^2 \|X_i\|_E^2 \). While, if out of this regime, then \( X \) is sub-exponential with squared norm \( \|X\|_E^2 = \sum_{i=1}^N a_i^2 \|X_i\|_E^2 \).
In the case that $\alpha_i = 1/N$, we have the Bernstein’s inequality that for all $t \geq 0$:

$$P \left( \frac{1}{N} \sum_{i=1}^{N} X_i \geq t \right) \leq 2 \exp \left( -\min \left\{ \frac{N^2t^2}{2 \sum_{i=1}^{N} \|X_i\|_E^2}, \frac{Nt}{\max_{1 \leq i \leq N} \|X_i\|_E} \right\} \right).$$

Finally, the relations between sub-Gaussian and sub-exponential variables are worth mentioning. We have seen above that a sub-Gaussian variable is itself sub-exponential by definition. Moreover, the product of two sub-Gaussian variables is sub-exponential. As a special case, a sub-Gaussian variable squared is sub-exponential, as in the case above that $Z \sim \mathcal{N}(0,1)$ is sub-Gaussian, and $X = Z^2 - 1$ is sub-exponential. Now we are ready to use the properties of such variables to establish fluctuation-response inequalities.

### III. APPLICATION IN NONLINEAR RESPONSE

#### A. General theory

We mainly follow Dechant and Sasa’s idea [18]. In the general setting, we have a physical system, characterized by a probability distribution $P \in \mathcal{P}$, where $\mathcal{P}$ is a probability family on a measurable space $\Omega$. We assume for simplicity here that for any $P \in \mathcal{P}$, $P$ has a density function with respect to some dominating measure. Such distributions can be those that characterize steady states or stochastic trajectories. For a random variable $X$, we denote its centered version as $\Delta X = X - \mathbb{E}X$. First let us use $P_0$ as the reference probability that describes the unperturbed state or the forward process, and $P_1$ as the distribution for the perturbed state or the backward process. One is interested how the change in distribution affects the ensemble average of $X$. Suppose for now $\mathbb{E}_0 X - \mathbb{E}_1 X \geq 0$, then starting from the moment generating function of $\Delta X$ with respect to $P_1$ (when well defined), we have for $s \geq 0$ that

$$\log \mathbb{E}_1 e^{s \Delta X} = \log \left( \int_{\Omega} e^{s \Delta X(\omega)} P_1(\omega) d\omega \right) = \log \left( \int_{\Omega} e^{s[X(\omega) - \mathbb{E}_1 X]} \frac{P_1(\omega)}{P_0(\omega)} P_0(\omega) d\omega \right) = \log \mathbb{E}_0 \left( e^{s[X - \mathbb{E}_1 X]} \frac{P_1}{P_0} \right) \geq \mathbb{E}_0 \log \left( e^{s[X - \mathbb{E}_1 X]} \frac{P_1}{P_0} \right) \text{ (by Jensen’s)} = \mathbb{E}_0 X - \mathbb{E}_1 X - D_{KL}(P_0 \parallel P_1),$$

where $D_{KL}(P_0 \parallel P_1) = -\int_{\Omega} \log(P_1/P_0) P_0 d\omega$ is defined to be the Kullback-Leibler divergence between $P_0$ and $P_1$. It is always nonnegative and equal to 0 only when $P_0 = P_1$.

Rearrange to obtain

$$\mathbb{E}_0 X - \mathbb{E}_1 X \leq \frac{1}{s} \left[ \log \mathbb{E}_1 e^{s \Delta X} + D_{KL}(P_0 \parallel P_1) \right]$$

for all $s \geq 0$ when $\mathbb{E}_1 e^{s \Delta X}$ is well defined. Similarly, if $\mathbb{E}_0 X - \mathbb{E}_1 X < 0$, then for $s \geq 0$ we have

$$\log \mathbb{E}_1 e^{-s \Delta X} \geq -s (\mathbb{E}_0 X - \mathbb{E}_1 X) - D_{KL}(P_0 \parallel P_1),$$

which implies

$$\mathbb{E}_1 X - \mathbb{E}_0 X \leq \frac{1}{s} \left[ \log \mathbb{E}_1 e^{-s \Delta X} + D_{KL}(P_0 \parallel P_1) \right]$$

for all $s \geq 0$ when $\mathbb{E}_1 e^{-s \Delta X}$ is well defined. Hence, combining these results, we have a Chernoff-like inequality that

$$\mathbb{E}_1 X - \mathbb{E}_0 X \leq \inf_{s \geq 0} \left\{ \frac{1}{s} \left[ \log \mathbb{E}_1 e^{s \Delta X} + D_{KL}(P_0 \parallel P_1) \right] \right\},$$

(7)

where $\xi = \text{sign}(\mathbb{E}_0 X - \mathbb{E}_1 X)$.

On the other hand, due to symmetry, it is straightforward to have that

$$\mathbb{E}_1 X - \mathbb{E}_0 X \leq \inf_{s \geq 0} \left\{ \frac{1}{s} \left[ \log \mathbb{E}_0 e^{s \Delta X} + D_{KL}(P_1 \parallel P_0) \right] \right\}. \quad (8)$$

Hence, one can take the minimum of these two upper bounds as $b = \min \{b_1, b_2\}$, where $b_1$ and $b_2$ denote the above bounds in (7) and (8). Note that neither of $b_1$ and $b_2$ is symmetric in $P_0$ and $P_1$. One might tend to construct symmetric bounds in the form $|\mathbb{E}_1 X - \mathbb{E}_0 X| \leq (b_1 + b_2)/2$, $|\mathbb{E}_1 X - \mathbb{E}_0 X| \leq \sqrt{b_1 b_2}$, etc, however, these bounds are less tight than $b$. In the following, without the loss of generality, we will assume (7) provides a tighter bound than (8) does. But before we show how the bound could be explicitly expressed in terms of sub-Gaussian or sub-exponential norm, other than of cumulants as in Ref. [18], let us consider two important situations where $D_{KL}$ can be (at least partially) expressed by thermodynamical quantities.

**Case I.** First, let us consider the case that $P_0$ and $P_1$ are the corresponding Boltzmann distributions at two states with Hamiltonians $H_0$ and $H_1$, respectively. Hence $P_0 = e^{-\beta H_0}/Z_0$ and $P_1 = e^{-\beta H_1}/Z_1$, where $\beta$ denotes the inverse temperature of the system (we set the Boltzmann constant $k_B = 1$) and $Z_{0,1}$ are partition functions. Then $D_{KL}(P_0 \parallel P_1)$ can be written as

$$D_{KL}(P_0 \parallel P_1) = \mathbb{E}_0 \log \left( \frac{e^{-\beta H_0}/Z_0}{e^{-\beta H_1}/Z_1} \right) = -\beta \mathbb{E}_0 (H_0 - H_1) + \log \left( \frac{Z_1}{Z_0} \right).$$

Note $-\log Z_{0,1}/\beta = F_{0,1}$ where $F_{0,1}$ are the Helmholtz free energies, and $\mathbb{E}_0 H_0$ is the internal energy $U_0$, hence $D_{KL}(P_0 \parallel P_1) = -\beta (U_0 - \mathbb{E}_0 H_1) = \beta (F_1 - F_0)$. The only term that has no direct thermodynamical correspondence is $\mathbb{E}_0 H_1$, which can be written as $\mathbb{E}_0 H_1 = \mathbb{E}_1 H_1 + (\mathbb{E}_0 H_1 - \mathbb{E}_1 H_1) =$.
When $P_0 = P_1$, the bound is tight. However, without more information, there is no guarantee on the tightness of this bound for general $P_0$ and $P_1$. Recently there have been some attempts to find the thermodynamical meaning of Wasserstein distance (22), however, in this work we will come back to Eq. (9) later using sub-Gaussian or sub-exponential norm.

The above argument can be substantially simplified in the linear response regime. If $H_1 = H_0 + \varepsilon A$, where $\varepsilon$ is a small parameter, then $-\beta E_{H_0} (H_1 - H_0) = \varepsilon \beta E_{0} A$, $Z_1 \approx Z_0 (1 - \varepsilon \beta E_0 A + \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A)$, and $\log (Z_1 / Z_0) \approx - \varepsilon \beta E_0 A + \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A$. Hence we have $D_{KL} (P_1 || P) = \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A$.

It is interesting to see that although in general $D_{KL} (P_0 || P) \neq D_{KL} (P_1 || P)$, in this linear region they are the same. Note $D_{KL} (P_1 || P_0) = -\beta E_{H_1} (H_1 - H_0) + \log (Z_1 / Z_0)$. We can see from above that $\log (Z_0 / Z_1) \approx \varepsilon \beta E_0 A - \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A$.

Also, for an arbitrary quantity $B$, we have $E_1 B \approx (1 + \frac{1}{2} \varepsilon \beta E_0 A) E_0 B e^{-\varepsilon \beta A} \approx E_0 B - \varepsilon (E_0 AB - E_0 A E_0 B)$, where $cov_0 (A, B)$ is the correlation between $A$ and $B$ under $P_0$. Let $A = B$, we have $-\beta E_1 (H_1 - H_0) = -\varepsilon \beta E_0 A \approx -\varepsilon \beta E_0 A + \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A$. So $D_{KL} (P_1 || P_0) \approx \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A$, and we conclude that in the linear region it holds

$$D_{KL} (P_0 || P_1) \approx D_{KL} (P_1 || P_0) \approx \frac{1}{2} \varepsilon^2 \beta^2 \var_0 A. \quad (11)$$

**Case II.** Second, as is well known, if $P_0$ denotes the probability for a forward path $\omega$, and $P_1$ is the backward probability for the time-reversed path $\omega^\dagger$, then $D_{KL} (P_1 || P_0)$ is nothing but the entropy production $\Delta S$;

$$D_{KL} (P_0 || P_1) = \Delta S. \quad (12)$$

It is interesting to see that this relation is obtained if we interpret $H_1$ in Eq. (9) as the time-reversed Hamiltonian, thus the integral $\int_{\Omega} H_1 (P_0 - \omega) d\omega$ vanishes if the system Hamiltonian is invariant under time reversal.

In the following, we will use the sub-Gaussian or sub-exponential property to deal with the first term $\log E_1 e^{\varepsilon \Delta X}$ in the upper bound (7).

**B. Sub-Gaussian regime**

If $P_0$ and $P_1$ are sub-Gaussian, then by (1), we have

$$\log E_1 e^{\varepsilon \Delta X} \leq \frac{1}{2} \sigma^2 (\varepsilon^2) \frac{1}{2} \sigma^2 s^2,$$

inserting which into (7), and noting (2), we have

$$\begin{align*}
\left| E_{1} X - E_{0} X \right| & \leq \inf_{\sigma > 0} \left\{ \frac{1}{2} \sigma^2 s + \frac{1}{s} \right\} \left( \sqrt{D_{KL} (P_0 || P_1)} \right) \\
& = \inf_{\sigma > 0} \sigma \sqrt{2 D_{KL} (P_0 || P_1)} \\
& = \left| \Delta X \right|_{1G} / \sqrt{2 D_{KL} (P_0 || P_1)}. \quad (13)
\end{align*}$$

where $\left| \Delta X \right|_{1G}$ is the sub-Gaussian norm of $\Delta X$ with respect to $P_1$. The inequality (13) provides a universal bound on the relative difference between means of $X$ in terms of the sub-Gaussian norm.

**Case I.** Now let us consider Eq. (9) again. If $\Delta H_1$ is sub-Gaussian, which could be possible for a system with bounded energy, then insert Eq. (9) into (13). Taking $X = H_1$, we have

$$\begin{align*}
\left| E_{1} H_{1} - E_{0} H_{1} \right| & \leq \left| \Delta H_{1} \right|_{1G} \sqrt{2 D_{KL} (P_0 || P_1)} \\
& = \left| \Delta H_{1} \right|_{1G} / \sqrt{2 D_{KL} (P_0 || P_1)}.
\end{align*}$$

Solving this inequality, we get bounds for $E_1 H_1 - E_0 H_1$ as:

$$\begin{align*}
& \beta \left| \Delta H_1 \right|_{1G}^2 \left( 1 + \frac{2 (S_1 - S_0)}{\left| \Delta H_1 \right|_{1G}^2} \right) \\
& \leq E_1 H_1 - E_0 H_1 \\
& \leq \beta \left| \Delta H_1 \right|_{1G}^2 \left( 1 + \frac{2 (S_1 - S_0)}{\left| \Delta H_1 \right|_{1G}^2} \right),
\end{align*}$$

which holds under the condition that

$$S_1 - S_0 \geq - \frac{1}{2} \beta^2 \left| \Delta H_1 \right|_{1G}^2. \quad (14)$$

There are several thermodynamical implications of this result. First, we establish a connection between the difference in the ensemble averages of a Hamiltonian in terms of the entropy difference at two states. Second, although (14) is a mathematical requirement, physically we know that this must hold, and this implies that there is a bound on the entropy change between two states, which is given by the property $\left| \Delta H_1 \right|_{1G}$. At the first sight, it does not seem to make sense since there is no information of the other state (state “0”) involved, however, note that we have made an assumption that it is (7) rather than (8) that gives the tighter bound. Hence the information of state “0” is used. Actually, (14) has a natural twin by switching the indices 0 and 1:

$$\left| S_1 - S_0 \right| \geq - \frac{1}{2} \beta^2 \left| \Delta H_0 \right|_{1G}^2.$$

If $S_1 > S_0$, then (14) is trivial, but (15) upper bounds $S_1 - S_0$; while if $S_0 > S_1$, then (15) is trivial, but (14) upper bounds
$S_0 - S_1$. Hence we can summarize that

$$|S_1 - S_0| \leq \max \left\{ \frac{1}{2} \beta^2 \| \Delta H_0 \|_{G_2}^2, \frac{1}{2} \beta^2 \| \Delta H_1 \|_{G_2}^2 \right\}. $$ (15)

This result is a most general one concerning entropy change in the sub-Gaussian regime.

One can see that (15) holds in the simpler linear response case where $H_1 = H_0 + \varepsilon A$. By the thermodynamical relation $S = -\beta \partial \log Z / \partial \beta - \log Z$, as well as results above, we can find $S_1 - S_0 = -\beta \partial \log (Z_1 / Z_0) - \log (Z_1 / Z_0) \approx 2\varepsilon \beta E_0 A \sim \mathcal{O}(\varepsilon^1)$, while $\| \Delta H \|_{G_2}$ becomes close to $\text{var}(\Delta H) = \beta^2 \log Z / \partial \beta^2$, and in any state it does not depend on $A$ to the order of $\mathcal{O}(\varepsilon^0)$. Hence (15) naturally holds.

For some other sub-Gaussian physical quantity $X$ in the linear response regime, and in particular for $X$ that can be well approximated by a Gaussian, we have $\| \Delta X \|_{G_2} \approx \text{var}(\Delta X)$. As also shown in Sec. IIIA, $|E_1 X - E_0 X| \approx \varepsilon \beta \text{cov}_0(X, A)$, and to the leading order, $\text{var}(\Delta X) = \text{var}_0(\Delta X)$. Hence, by Eqs. (11) and (13) we obtain

$$|\text{cov}_0(X, A)| \leq \sqrt{\text{var}_0(\Delta X)} \sqrt{\text{var}_0 A} \leq \frac{1}{\sqrt{\text{var}_0 A}} \sqrt{\text{var}_0 A},$$

which shows that (13) coincides with the Cauchy-Schwarz inequality.

**Case II.** While for the path thermodynamics example (12), we have

$$|E_1 X - E_0 X| \leq \| \Delta X \|_{G_2} / \sqrt{2\Delta S}.$$

If $X$ changes sign under time reversal, then we further have

$$2(E X)^2 \leq \| \Delta X \|_{G_2}^2 \Delta S.$$ (16)

Note there is no need to specify with respect to which distribution the mean and sub-Gaussian norm are taken due to the time reversal operation. Again, when the Gaussian approximation is valid, we formally recover the thermodynamic uncertainty relation that [13, 20]

$$2(E X)^2 \leq \text{var}(X) \Delta S.$$

However, this inequality is derived based on a specific model, which although does represent a wide class of dynamics, is not as general as the setting in this work. Our theoretical results [13] and [16] hold in the absence of information about the detailed dynamics, thus substantially pushing existing bounds to the sub-Gaussian nonlinear regime, while at the expense of bound tightness.

**Operational issues.** Given above are mainly the theoretical results obtained by introducing sub-Gaussian variables. Now let us consider the operational part. Suppose we have $N$ independent, identically distributed data points $X_1, \ldots, X_N$. By the law of large numbers, $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i / N$ converges to $E X$. However, this process may need a huge number of data especially when the underlying distribution is non-Gaussian. Here we address the non-asymptotic bound that controls the error incurred by using $\hat{\mu}$ in our inequalities, for experimental inspection purposes. That is, if we use $\hat{\mu}_1$ and $\hat{\mu}_0$ in (13), what will result? Can we have some sense about the probability $P(|\hat{\mu}_1 - \hat{\mu}_0| - (E_1 X - E_0 X)) \geq t)$ for $t \geq 0$? Note $E(\hat{\mu} - E X) = 0$, hence $\hat{\mu} - E X$ is a centered variable. In fact, $\hat{\mu} - E X = \frac{1}{N} \sum_{i=1}^{N} (X_i - E X) = \frac{1}{N} \sum_{i=1}^{N} \Delta X_i$, and

$$E e^{s(\hat{\mu} - E X)} = E e^{\frac{1}{N} \sum_{i=1}^{N} \Delta X_i} = \prod_{i=1}^{N} E e^{\frac{1}{N} \Delta X_i}.$$

Since $\Delta X_i$ is a centered sub-Gaussian variable, then $E e^{s(\hat{\mu} - E X)} \leq e^{2s^2 \Delta X_i^2 / 2N^2}$, and we have

$$E e^{s(\hat{\mu} - E X)} \leq e^{2s^2 \Delta X_0^2 / 2N},$$

hence $\hat{\Delta} \mu \equiv \hat{\mu} - E X$ is a centered sub-Gaussian variable with norm $2(\| \Delta X \|_{G_2}^2 / \sqrt{N})$. And $P(|\hat{\Delta} \mu_1 - \hat{\Delta} \mu_0| - (E_1 X - E_0 X)) \geq t) = P(|\hat{\Delta} \mu_1 - \hat{\Delta} \mu_0| \geq t)$, then for all $s \in \mathbb{R}$,

$$E e^{s(\hat{\mu}_1 - \hat{\mu}_0)} = E e^{s \hat{\mu}_1} E e^{-s \hat{\mu}_0} \leq e^{s^2 \| \Delta X \|_{G_2}^2 / 2N} e^{s^2 \| \Delta X \|_{G_2}^2 / 2N},$$

hence $\hat{\Delta} \mu_1 \equiv \hat{\mu}_1 - \hat{\mu}_0$ is also sub-Gaussian, with norm $\sqrt{(\| \Delta X \|_{G_2}^2 + \| \Delta X \|_{G_2}^2) / N}$. This leads to the concentration inequality that

$$P(|\hat{\Delta} \mu_1 - \hat{\mu}_0| - (E_1 X - E_0 X)) \geq t) = P(|\hat{\Delta} \mu_1 - \hat{\mu}_0| \geq t) \leq 2 \exp \left( -\frac{N t^2}{2(\| \Delta X \|_{G_2}^2 + \| \Delta X \|_{G_2}^2) / N} \right),$$

which is indeed an application of the Hoeffding bound (3). Therefore, for all $\delta \in (0, 1)$, we have probability at least $1 - \delta$ that

$$|\hat{\Delta} \mu_1 - \hat{\mu}_0| \leq \sqrt{\frac{2(\| \Delta X \|_{G_2}^2 + \| \Delta X \|_{G_2}^2) \log \left( \frac{2}{\delta} \right)}{N}}.$$ (17)

Put it another way, one has probability at least $1 - \delta$ to have a desired error bound $\varepsilon$ on the difference between $\hat{\mu}_1 - \hat{\mu}_0$ and $E_1 X - E_0 X$, if the number of data points $N$ is on the order of $2(\| \Delta X \|_{G_2}^2 + \| \Delta X \|_{G_2}^2) \log(2/d)/\varepsilon^2$.

Also, since $|\hat{\mu}_1 - \hat{\mu}_0| \leq |\Delta \mu_1 | - | \Delta \mu_0 | + | E_1 X - E_0 X |$, then by (13), $|\hat{\mu}_1 - \hat{\mu}_0 | \leq | \Delta \mu_1 - \Delta \mu_0 | + | \Delta X |_{G_2} \sqrt{2D_{KL}(P_0 / P_1)}$. Thus (17) suggests that with probability more than $1 - \delta$, we have

$$|\hat{\mu}_1 - \hat{\mu}_0 | \leq \sqrt{\frac{2(\| \Delta X \|_{G_2}^2 + \| \Delta X \|_{G_2}^2) \log \left( \frac{2}{\delta} \right)}{N}} + | \Delta X |_{G_2} \sqrt{2D_{KL}(P_0 / P_1)},$$

which sets an upper bound for the absolute difference in sample means at different states. If there is some way to effectively estimate the norm and the Kullback-Leibler divergence (or their upper bounds), these results could be applied then.
C. Sub-exponential regime

In this case, one might directly insert (4) and (5) into (7) to get
\[ |E_1 X - E_0 X| \leq \|\Delta X\|_{1E} \sqrt{2D_{KL}(P_0 \| P_1)}, \tag{18} \]
which is formally almost identical to (13), with only the sub-exponential norm replacing the sub-Gaussian norm. However, there is a tacit constraint on (18). Note by definition that \(|s| \leq c_{1E}/\|\Delta X\|_{1E}\), the global infimum in (7) is achieved when \(s\sqrt{\|\Delta X\|^2_{1E}/2} = D_{KL}(P_0 \| P_1)/s\), i.e., \(s^2 = 2D_{KL}(P_0 \| P_1)/\|\Delta X\|^2_{1E}\). If this can be satisfied, then by the definition (4), we must have
\[ D_{KL}(P_0 \| P_1) \leq c_{1E}^2/2. \tag{19} \]
If this is the case, then everything we do to sub-Gaussian variables is the same here. However, this condition is somewhat demanding in general, and if it cannot be satisfied, then the infimum is obtained on the boundary \(s = c_{1E}/\|\Delta X\|_{1E}\), hence
\[ |E_1 X - E_0 X| \leq \|\Delta X\|_{1E} \left(\frac{c_{1E}}{2} + \frac{D_{KL}(P_0 \| P_1)}{c_{1E}}\right). \tag{20} \]
One might be tempted to choose a \(c_{1E}\) in our definition (4) so that the condition on \(D_{KL}(P_0 \| P_1)\) can be easier to be satisfied, however, a bigger \(c_{1E}\) may also result in a bigger \(\|\Delta X\|_{1E}\), and consequently a less tight bound. Without further information of the dynamics, there seems no reason for a natural optimal choice of \(c_{1E}\).

Case I. Let us also take \(X = H_1\). For physical systems, it is also common to see energy distribution in the form \(e^{-\beta E}\), which is actually the Gamma distribution and falls in the class of sub-exponential distribution. Hence it is also possible that \(H_1\) is sub-exponential. Apply the similar analysis as in the sub-Gaussian case, we obtain the bound on the entropy change that
\[ S_1 - S_0 \geq \max\left\{\left(\frac{c_{1E}}{\|\Delta H_1\|_{1E}} - \beta\right)(E_1 H_1 - E_0 H_1) - \frac{c_{1E}^2}{2}, \right. \]
\[ \left. - \left(\frac{c_{1E}}{\|\Delta H_1\|_{1E}} + \beta\right)(E_1 H_1 - E_0 H_1) + \frac{c_{1E}^2}{2}\right\}, \tag{21} \]
which is one step ahead of the previous bound (9). It is also possible to bound \(|S_1 - S_0|\), but we omit the result here.

Case II. Next, the thermodynamic uncertainty relation can also be addressed in the sub-exponential situation. Again, if the condition (19) is satisfied, then formally it is straightforward to have for a time-antisymmetric quantity \(X\) that
\[ 2(\langle X \rangle)^2 \leq \|\Delta X\|_{1E}^2 \Delta S. \tag{22} \]
And when the condition is not satisfied,
\[ |E X| \leq \|\Delta X\|^2_{1E} \left(\frac{c_{1E}}{2} + \frac{\Delta S}{c_{1E}}\right). \tag{23} \]

Operational issues. Finally, operationally if \(X\) is measurable at different states, then the error incurred by using the sample mean instead of the ensemble average can also be quantified. Following almost exactly the same steps as shown in the sub-Gaussian case, we can establish the concentration bound on such an error. Suppose we have \(N\) independent measurements at each state, and let the corresponding sample mean be \(\hat{\mu} = \sum_{i=1}^{N} X_i/N\). Applying the Bernstein bound (9), we have
\[ P\left(|\hat{\mu} - \mu_0| - (E_1 X - E_0 X) \geq t\right) \]
\[ \leq 2 \exp\left(\frac{-\min\left\{2\|\Delta X\|^2_{1E} + \|\Delta X\|^2_{0E}\right\}}{2Nt} \right) \frac{\left[2Nt \max\{\|\Delta X\|_{1E}, \|\Delta X\|_{0E}\}\right]}{\log(2/\delta)/2N + \log(2/\delta)/4N}, \tag{24} \]
Note that \(\min\{b_1, b_2\} \leq (b_1 + b_2)/2\), hence for all \(0 < \delta < 1\), we have the non-asymptotic bound that with probability more than \(1 - \delta\), the difference between \(\hat{\mu}_0 - \mu_0\) and \(E_1 X - E_0 X\) is no more than \(\sqrt{\|\Delta X\|^2_{1E} + \|\Delta X\|^2_{0E}}\log(2/\delta)/2N + \max\{\|\Delta X\|_{1E}, \|\Delta X\|_{0E}\}\log(2/\delta)/4N\), and the concentration bound for \(\hat{\mu}_0 - \mu_0\) follows as in the sub-Gaussian case. If the sub-exponential forms can be obtained or nontrivially upper bounded theoretically, then such inequalities can be used to evaluate experimental data.

IV. CONCLUSION AND DISCUSSION

In this work, we have introduced the concepts of sub-Gaussian and sub-exponential distributions, which seem less known to the statistical physics community. The motivation of our work is that sub-Gaussian and sub-exponential distributions, as natural generalizations to the Gaussian distribution that facilitates the linear response theory, seem to be particularly relevant to nonlinear response, as hinted by previous experimental and numerical findings.

Based on the sub-Gaussian or sub-exponential norm of a physical quantity, we are able to further develop the theory established by Dechant and Sasa [18] for the fluctuation-response relation in general situations out of equilibrium. We refine the bound for the difference in expected values (with respect to two distributions) of an arbitrary variable that falls within the sub-Gaussian or sub-exponential class. When the distributions considered are about two equilibrium states connected by an external perturbation, we also find a bound that links the entropy difference with the Hamiltonian fluctuation. When the distributions are interpreted as regards to the forward and backward processes, respectively, we obtain a generalized version of the thermodynamic uncertainty relation in each regime. Our results provide universal constraints on the thermodynamical processes without requiring more detailed information of the system in question. But surely as such information is available, more accurate results with tighter bounds are expected.

Finally, it will be interesting to experimentally test some of our results in real physical systems, which include not only the bounds, but also the plausible transition path from the
Gaussian to sub-Gaussian and sub-exponential distributions as some control parameter is varied. We have also provided error bounds for the estimation of the difference in expected values in either sub-Gaussian or sub-exponential cases. A practical challenge, however, is the estimation of the sub-Gaussian or sub-exponential norm itself. To our knowledge, there does not exist an effective way to perform a point estimation of such norms with a quantitative margin of error. Nonetheless, one can instead use an upper bound of the norm to test the results, which may be easier to get access to theoretically or experimentally, but at the expense of a less tight bound.

[1] L. D. Landau, and E. M. Lifshitz, *Statistical Physics, Part I* (Pergamon Press, New York, 1980), 3rd ed.
[2] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer, New York, 1998), 2nd ed.
[3] U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Phys. Rep. **461**, 111 (2008).
[4] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, Phys. Rev. Lett. **78**, 2690 (1997).
[5] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).
[6] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamante, Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski’s Equality, Science **296**, 1832 (2002).
[7] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., and C. Bustamante, Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies, Nature (London) **437**, 231 (2005).
[8] B. Cleuren, C. Van den Broeck, and R. Kawai, Fluctuation and Dissipation of Work in a Joule Experiment, Phys. Rev. Lett. **96**, 050601 (2006).
[9] V. Blickle, T. Speck, L. Helden, U. Seifert, and C. Bechinger, Thermodynamics of a Colloidal Particle in a Time-Dependent Nonharmonic Potential, Phys. Rev. Lett. **96**, 070603 (2006).
[10] M. Bonaldi et al., Nonequilibrium Steady-State Fluctuations in Actively Cooled Resonators, Phys. Rev. Lett. **103**, 010601 (2009).
[11] A. Engel, Asymptotics of work distributions in nonequilibrium systems, Phys. Rev. E **80**, 021120 (2009).
[12] P. Chvosta, D. Lips, V. Holubec, A. Ryabov, and P. Maass, Statistics of work performed by optical tweezers with general time-variation of their stiffness, arXiv preprint arXiv:2003.07875v1 (2020).
[13] H. J. D. Miller, M. Scandi, J. Anders, and M. Perarnau-Llobet, Work Fluctuations in Slow Processes: Quantum Signatures and Optimal Control, Phys. Rev. Lett. **123**, 230603 (2019).
[14] Z. Fei and H. T. Quan, Nonequilibrium Green’s function’s approach to the calculation of work statistics, arXiv preprint arXiv:2001.08061v1 (2020).
[15] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint* (Cambridge University Press, New York, 2019).
[16] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science* (Cambridge University Press, New York, 2018).
[17] \[ P(X \leq -x) = \int_{-\infty}^{-x} P(u)du = \int_{x}^{\infty} P(u)e^{-u}du \leq e^{-x}\int_{-\infty}^{\infty} P(u)du = e^{-x}. \]
[18] A. Dechant and S.-i. Sasa, Fluctuation-response inequality out of equilibrium, Proc. Natl. Acad. Sci. U.S.A. **117**, 6430 (2020).
[19] A. C. Barato and U. Seifert, Thermodynamic Uncertainty Relation for Biomolecular Processes, Phys. Rev. Lett. **114**, 158101 (2015).
[20] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. England, Dissipation Bounds All Steady-State Current Fluctuations, Phys. Rev. Lett. **116**, 120601 (2016).
[21] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, 2009).
[22] A. Dechant and Y. Sakurai, Thermodynamic interpretation of Wasserstein distance, arXiv preprint arXiv:1912.08405v1 (2019).
[23] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Dissipation: The Phase-Space Perspective, Phys. Rev. Lett. **98**, 080602 (2007).