Information Inequality Interpretation of Uncertainty Relation in Stochastic Process

Yoshihiko Hasegawa and Tan Van Vu
Department of Information and Communication Engineering,
Graduate School of Information Science and Technology,
The University of Tokyo, Tokyo 113-8656, Japan
(Dated: September 11, 2018)

Thermodynamic uncertainty relation is an inequality for precision of thermodynamic quantities, stating that it is impossible to attain higher precision than the bound defined by the entropy production. Such trade-off relations between “cost” and “quality” exist in several fields. Particularly, in statistical inference theory, the information inequalities assert that it is infeasible for any statistical estimators to achieve smaller error than the prescribed bound. Inspired by the similarity between the thermodynamic uncertainty relation and the information inequalities, we apply the information inequalities to systems described by Langevin equations and derive the bound for the variance of thermodynamic quantities. When applying the Cramér–Rao inequality, the obtained inequality is shown to reduce to the fluctuation-response inequality, which relates the fluctuation with the linear response of systems against perturbations. We also apply Hammersley–Chapman–Robbins inequality to the systems and find a relation, giving the lower bound of the ratio between the variance and the sensitivity of the systems in response to arbitrary perturbations. To confirm the second inequality, we apply it to a stochastic limit cycle oscillator and numerically show that the ratio between the phase variance and the phase sensitivity is bounded from below by the reciprocal of the Pearson divergence.

Introduction.—In these two decades, by virtue of advancement of stochastic thermodynamics, substantial progresses have been made in universal relations among thermodynamic quantities, such as fluctuation theorems and generalized second laws [1–4]. One of the fundamental achievements is the thermodynamic uncertainty relation (TUR) [5–14], which states that fluctuation of thermodynamic quantities (e.g., current) is bounded from below by the reciprocal of the entropy production. Such trade-off relations between “cost” and “quality” exist in several fields. Particularly, in statistical inference theory, the information inequalities assert that it is infeasible for any statistical estimators to achieve smaller error than the prescribed bound. Inspired by the similarity between the thermodynamic uncertainty relation and the information inequalities, we apply the information inequalities to systems described by Langevin equations and derive the bound for the variance of thermodynamic quantities.

The obtained relation reduces to a recently discovered fluctuation-response inequality (FRI) [15], which provides an inequality between fluctuations and linear response of the system. Using the equality condition of CRI, we identify the necessary and sufficient condition for one-dimensional Langevin equation to attain the equality of TUR. Furthermore, we apply Hammersley–Chapman–Robbins inequality (HCRI), which is a generalization of CRI, to the systems to show that, for any perturbation, the ratio between the variance and the sensitivity is bounded from below by the Pearson divergence. Although abovementioned CRI and FRI only hold for weakly perturbed cases, the HCRI-based relation is general and valid even for out of linear response regime. As an application of the HCRI-based relation, we obtain an explicit inequality between the phase variance and the phase sensitivity of stochastic limit cycle oscillators, which were empirically known to be trade-off factors. Our Letter shows that theories of statistical inference provide new insight into uncertainty relations in stochastic thermodynamic systems.

Model.—We consider the following N-dimensional Langevin equation for $\mathbf{x} \equiv [x_1, x_2, \ldots, x_N]^T$:

$$\dot{\mathbf{x}} = \mathbf{A}_\theta(\mathbf{x}, t) + \sqrt{2} \mathbf{C}(\mathbf{x}, t) \mathbf{\xi}(t), \quad (1)$$

where $\mathbf{\xi}(t) \equiv [\xi_1(t), \xi_2(t), \ldots, \xi_N(t)]^T \in \mathbb{R}^{N \times 1}$ is the white Gaussian noise with $\langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \delta(t - t')$, $\mathbf{A}_\theta(\mathbf{x}, t) \in \mathbb{R}^{N \times N}$ is a drift vector with a parameter $\theta \in \mathbb{R}$, and $\mathbf{C}(\mathbf{x}, t) \in \mathbb{R}^{N \times N}$ is a noise matrix. We employ Ito stochastic integration for Eq. (1). $\theta$ is a parameter which is to be estimated by some estimators. We assume that $\theta$ is a scalar for simplicity, but we can easily generalize to a multidimensional vector $\theta$. Let $\mathbf{P}_\theta(\mathbf{x}, t)$ be a probability density function of $\mathbf{x}$ at time $t$. Defining $[B_{ij}(\mathbf{x}, t)] = \mathbf{B}(\mathbf{x}, t) \equiv \mathbf{C}(\mathbf{x}, t) \mathbf{C}(\mathbf{x}, t)^T$, a Fokker–Planck equation (FPE) of Eq. (1) reads [16, 17] $\frac{\partial}{\partial t} \mathbf{P}_\theta(\mathbf{x}, t) = \hat{\mathbf{L}}_\theta(\mathbf{x}, t) \mathbf{P}_\theta(\mathbf{x}, t)$ where $\hat{\mathbf{L}}_\theta \equiv -\sum_i \partial_{x_i} \mathbf{A}_{\theta,i}(\mathbf{x}, t) + \sum_{i,j} \partial_{x_i} \partial_{x_j} B_{ij}(\mathbf{x}, t)$ is an
FPE operator. The probability current is \( J_{\theta,i}(x,t) = \left\{ A_{\theta,i}(x,t) - \sum_j \partial_{x_j} B_{ij}(x,t) \right\} P_\theta(x,t) \).

Suppose we want to estimate the parameter \( \theta \) from a measurement of the Langevin equation. The measurement is a stochastic trajectory within an interval from \( t = 0 \) to \( T \) generated by Eq. (2). We discretize time by dividing the interval \([0, T]\) into \( K\) equi-partition intervals with time resolution \( \Delta t \), where \( T = K \Delta t \), \( t^k = k \Delta t \), and \( x^k \equiv x(t^k) \) (we use superscripts for specifying a point in temporal and subscript for an element of vectors or matrices). Let \( \Gamma \equiv \{ x(t) \}_{t=0}^T \) be a stochastic realization of Eq. (21) and \( P_\theta(\Gamma|x^0) \) be likelihood given \( x^0 \) at \( t = 0 \). By using a path-integral representation, \( P_\theta(\Gamma|x^0) \) is given by [14][18][20]

\[
    P_\theta(\Gamma|x^0) = \mathcal{N} \exp \left[ -\int_0^T dt \, \mathcal{A}_\theta(x(t), \dot{x}(t)) \right], \tag{2}
\]

where \( \mathcal{A}_\theta(x, \dot{x}) \equiv \frac{1}{4} \left\{ (\dot{x} - A_\theta)^T B^{-1} (\dot{x} - A_\theta) \right\} \) is a stochastic action and \( \mathcal{N} \) is a term that does not depend on \( \theta \). A path-integral representation of Eq. (2) employs a pre-point discretization. Due to the pre-point discretization, cross terms such as \( \int A_\theta(x, t) B^{-1}(x, t) \dot{x} dt \) should be interpreted \( \int A_\theta(x, t) B^{-1}(x, t) \dot{x} dt \) where \( \cdot \) denotes the Ito product [21]. Although the action \( \mathcal{A}_\theta \) appears to be different for a mid-point discretization [22], we can show that \( \mathcal{A}_\theta \) of the mid-point discretization reduces to that of the pre-point discretization for additive noise systems [21][23]. The unconditional likelihood is \( P_\theta(\Gamma) = P_\theta(\Gamma|x^0) P_\theta(x^0) \) where \( P_\theta(x^0) \) is initial probability density of \( x^0 \) at \( t = 0 \). The log-likelihood is \( \ln P_\theta(\Gamma) = \ln P_\theta(x^0) - \frac{1}{4} \int_0^T dt \, \left( \dot{x} - A_\theta \right)^T B^{-1} \left( \dot{x} - A_\theta \right) \) where we dropped \( \ln \mathcal{N} \) since it does not depend on \( \theta \).

Let us consider an estimator \( \Theta(\Gamma) \), which is an arbitrary function of a trajectory \( \Gamma \) (since \( \Gamma \) is a stochastic trajectory, \( \Theta(\Gamma) \) is a random variable). For \( f(\Gamma) \) which is an arbitrary function of \( \Gamma \), let \( \langle f(\Gamma) \rangle_\theta \equiv \int \mathcal{D}\Gamma \, f(\Gamma) P_\theta(\Gamma) \). Since \( \langle \Theta(\Gamma) \rangle_\theta \) contains information of \( \theta \), we may write \( \langle \Theta(\Gamma) \rangle_\theta = \psi(\theta) \). In this way, \( \Theta(\Gamma) \) can be regarded as an estimator of \( \psi(\theta) \). In statistical inference, CRI is the fundamental relation which gives the lower bound of the variance of estimators. By using CRI [24][26] to \( \Theta(\Gamma) \), we obtain the following inequality [21]

\[
    \langle \partial_\theta \langle \Theta(\Gamma) \rangle_\theta \rangle_\theta^2 \geq \frac{1}{\mathcal{I}(\theta)}, \tag{3}
\]

\[
    \mathcal{I}(\theta) = -\langle \partial_\theta \ln P_\theta(x_0) \rangle_\theta + \frac{1}{2} \int_0^T dt \, \langle \partial_\theta A_\theta^T \rangle_\theta B^{-1} \langle \partial_\theta A_\theta \rangle_\theta, \tag{4}
\]

where \( \text{Var}_\theta[f(\Gamma)] \equiv \left\{ \langle f(\Gamma) \rangle_\theta - \langle \langle f(\Gamma) \rangle_\theta \rangle_\theta \right\}^2 \) and \( \mathcal{I}(\theta) \equiv -\langle \partial_\theta \ln P_\theta(\Gamma) \rangle_\theta \) is the Fisher information [21]. From the equality condition of CRI, the equality of Eq. (3) is satisfied if and only if \( \Theta(\Gamma) \) is expressed as

\[
    \partial_\theta \ln P_\theta(\Gamma) = \mu(\theta) [\Theta(\Gamma) - \psi(\theta)], \tag{5}
\]

where \( \mu(\theta) \) is an arbitrary scaling function. Equation (5) is a direct consequence of the equality condition of Cauchy–Schwarz inequality.

CRI of Eq. (3) reduces to recently proposed FRI [13]. Suppose \( A_\theta(x, t) = A(x, t) + \theta \mathcal{Y}(x, t) \), where \( \theta \in \mathbb{R} \) is a sufficiently small parameter and \( \mathcal{Y}(x, t) \in \mathbb{R}^{N \times 1} \) is an arbitrary perturbation. Since \( \theta \) is assumed to be sufficiently small, the sensitivity \( \partial_\theta \langle \Theta(\Gamma) \rangle_\theta \) can be approximated by \( \partial_\theta \langle \Theta(\Gamma) \rangle_\theta \theta \equiv \langle \langle \Theta(\Gamma) \rangle \rangle_\theta - \langle \langle \Theta(\Gamma) \rangle \rangle_\theta \rangle_\theta \). Plugging these expressions into Eq. (4), we obtain FRI:

\[
    \text{Var}_\theta \langle \Theta(\Gamma) \rangle_\theta \langle \Theta(\Gamma) \rangle_\theta - \langle \langle \Theta(\Gamma) \rangle \rangle_\theta \langle \langle \Theta(\Gamma) \rangle \rangle_\theta \rangle_\theta \end{array} \geq \mathcal{C}^{-1} \end{array}, \tag{6}
\]

where \( \mathcal{C} = \sqrt{\mathcal{I}(\theta)} \) is an inequality for the variance of estimators and is a generalization of CRI. By using HCRI to the Langevin equations [Eq. (1)], we obtain the following relation [21]:

\[
    \frac{\text{Var}_\theta \langle \Theta(\Gamma) \rangle_\theta^2}{\langle \langle \Theta(\Gamma) \rangle_\theta - \langle \langle \Theta(\Gamma) \rangle_\theta \rangle_\theta \rangle_\theta^2} \geq \frac{1}{\text{D}P_\theta[P_\theta|\theta]}, \tag{6}
\]

where \( \text{D}P_\theta[P_\theta|\theta] \) is the Pearson divergence between \( P_\theta \) and \( P_\theta \) defined by \( \text{D}P_\theta[P_\theta|\theta] \equiv \int \mathcal{D}\Gamma \left( \frac{P_\theta(\Gamma)}{P_\theta(\Gamma)} - 1 \right)^2 P_\theta(\Gamma) \), and \( \theta \in \mathbb{R} \) is an arbitrary parameter that satisfies \( \theta \neq \theta \). Equation (6) is the main result of this Letter. \( \langle \langle \Theta(\Gamma) \rangle_\theta - \langle \langle \Theta(\Gamma) \rangle_\theta \rangle_\theta \rangle_\theta \) describes the difference between two dynamics characterized by \( \theta \) and \( \theta \), which is the sensitivity of the system. Thus the ratio between variance of unperturbed dynamics and the sensitivity is bounded by the reciprocal of the Pearson divergence between the two dynamics. For \( \theta = \theta + \Delta \theta \), Eq. (6) reduces to CRI [Eq. (3)] and FRI, and thus HCRI is a generalization of CRI and FRI [21]. Although CRI and FRI only hold locally around some parameter \( \theta \) and thus they hold for sufficiently weak perturbation, Eq. (6) can be used for out of linear response regime. Although Kullback–Leibler (KL) divergence plays important roles in stochastic thermodynamics [24][25] and the Pearson divergence is a fundamental measure in machine learning [21], the Pearson divergence rarely appears in the context of stochastic thermodynamics.

**CRI in current TUR.** We show statistical inference view of the current TUR. Reference [15] re-derived the finite-time current TUR [10] with FRI using a notion of virtual perturbation. Let us consider a Langevin equation \( \dot{x} = A(x) + \sqrt{2C(x)} \xi(t) \) [cf Eq. (21)]. The current TUR considers the generalized current \( \Theta_{\text{cur}}(\Gamma) \equiv \int_0^T dt \, \langle \partial_\theta \ln P_\theta(\Gamma) \rangle_\theta \).
\[ \int_0^T \Lambda(x)^T \circ \dot{x} \, dt \] where \( \Lambda(x) \in \mathbb{R}^{N \times 1} \) is an arbitrary projection function and the product \( \circ \) should be interpreted in the Stratonovich sense. In their derivation of TUR \([13]\), they considered a modified drift given by 
\[ A_\theta(x) = (\theta + 1)A_i(x) - P^{ss}(x)^{-1} \theta \sum_j \partial x_j B_{ij}(x) P^{ss}(x), \]
where \( P^{ss}(x) \) is the steady-state distribution of the unperturbed dynamics (i.e., dynamics of \( \theta = 0 \) case). Furthermore, \( \mathcal{I}(0) \), which is a denominator of the lower bound of FRI, turned out to be the entropy production. Therefore, from a statistical inference viewpoint, the current \( \Theta_{\text{cri}}(\Gamma) \) is an estimator which infers \( \theta \), and the entropy production corresponds to the Fisher information in \( \theta \)-space. The Fisher information describes log likelihood change when varying a parameter \( \theta \). If the change is large, the curvature of the log likelihood becomes steeper, which makes the parameter inference more accurate.

The estimator which attains the CRI bound is known as an efficient estimator. When the equality condition of CRI is satisfied, then that of TUR also holds. The necessary and sufficient condition for \( \Theta_{\text{cri}}(\Gamma) \) to become an efficient estimator is that the following relation holds [Eq. (5)]:
\[ \partial \ln \mathcal{P}_\theta(\Gamma) \propto \Theta_{\text{cri}}(\Gamma) - \psi(\theta). \]
We differentiate \( \ln \mathcal{P}_\theta(\Gamma) \) to obtain
\[ \partial \ln \mathcal{P}_\theta(\Gamma) \propto \int_0^T \left[ (J^{ss})^T B^{-1} \hat{x} - (J^{ss})^T B^{-1} A_\theta \right] \, dt, \]
where \( \hat{x} \) denotes the Ito product, and \( J^{ss}(x) \) is the probability current at steady-state. In Eq. (7), the first term is an Ito-type current and \( (J^{ss})^T B^{-1} / P^{ss} \) can be identified as the conjugate thermodynamic force \([8]\). Although we can convert the Ito-type current into its corresponding Stratonovich one \([21]\), we cannot still expect that the remaining terms reduce to \( \psi(\theta) \) in general. This shows that Stratonovich-type current which attains the equality condition of the current TUR does not exist for arbitrary \( A(x) \) and \( B(x) \). Still, the equality of TUR can be achieved for some \( A(x), B(x) \), and \( \Lambda(x) \). For one-dimensional system with periodic boundary conditions, using Eq. (7), we can obtain the necessary and sufficient condition for \( A(x), B(x), \) and \( \Lambda(x) \) to attain the equality \([21]\):
\[ A(x) = \kappa_1 \sqrt{B(x)} + \partial x B(x)/2, \]
\[ \Lambda(x) = \kappa_2 / \sqrt{B(x)}, \]
where \( \kappa_1 \) and \( \kappa_2 \) are arbitrary parameters. If and only if systems are represented by Eqs. (8) and (9), the equality of the current TUR is attained.

**HCRI in Ornstein-Uhlenbeck process.**— Next, we study HCRI of \([21]\) in a Langevin equation. We consider
\[ A_\theta(x, t) = -\partial_x V(x) + \theta u(t), \quad C(x, t) = \sqrt{D}, \]
in Eq. (10), where \( V(x) \) is a potential function and \( u(t) \) is an arbitrary input function. We consider the Ornstein-Uhlenbeck (OU) process, where \( V(x) = \alpha x^2/2 \) with \( \alpha > 0 \). An initial condition is \( x = 0 \) at \( t = 0 \). The Pearson divergence between \( \mathcal{P}_\theta(\Gamma) \) (perturbed) and \( \mathcal{P}_{\theta=0}(\Gamma) \) (unperturbed) is represented analytically by \([21]\)
\[ D_{\text{PE}}[\mathcal{P}_\theta||\mathcal{P}_{\theta=0}] = -1 + \exp \left( \frac{\theta^2}{2D} \int_0^T u(t)^2 \, dt \right). \]

When we define \( \Theta_x(\Gamma) \equiv \int_0^T \dot{x} \, dt \), \( \Theta_x(\Gamma) \) simply gives the position \( x(T) \). Therefore, HCRI of Eq. (6) reads \( \Psi_x / \mathcal{S}_x \geq 1 / D_{\text{PE}}[\mathcal{P}_\theta||\mathcal{P}_{\theta=0}] \) which \( \Psi_x \equiv \text{Var}_{\theta=0}[x(T)] \) and \( \mathcal{S}_x \equiv \langle [(x(T) - x(T)_{\theta=0})] \rangle \). We also consider the lower bound of FRI: \( 1 / [\theta^2 \mathcal{I}(0)] = 1 / \left[ \theta^2 \int_0^T u(t)^2 \, dt \right] \).

For the OU process with an arbitrary \( u(t) \geq 0 \) and the initial condition, we can show \( \Psi_x / \mathcal{S}_x \geq 1 / \theta^2 \mathcal{I}(0) \geq D_{\text{PE}}[\mathcal{P}_\theta||\mathcal{P}_{\theta=0}] \),

which indicates that the bound of FRI is always tighter than that of HCRI, and FRI always holds for the OU process.

**HCRI in double-well potential.**— We consider a strongly nonlinear double-well potential \( V(x) = -x^2/2 + x^4/4 + \beta x \), where \( \beta > 0 \) is a bias parameter, in Eq. (10). For \( \beta = 0 \), there are two stable points at \( x = \pm 1 \), and the left well becomes globally stable for \( \beta > 0 \). An initial condition is set to \( x = -1 \) at \( t = 0 \). Suppose \( u(t) = 1 \) for \( t > 0 \) and \( u(t) = 0 \) for \( t \leq 0 \). When \( \theta = 2\beta \), the globally stable well switches from left to right at \( t > 0 \) in the presence of the perturbation \( \theta u(t) \). We test HCRI for \( \Theta_x(\Gamma) \equiv \int_0^T \dot{x} \, dt \) in the double-well potential with Monte Carlo simulations. We randomly sample parameters \( \beta, \theta, D, \) and \( T \) (parameter ranges are shown in the caption of Fig. (11)), and solve the Langevin equation \( N^2 = 1.0 \times 10^6 \) times with time resolution \( h = 0.0002 \) for each of the selected parameter settings. In Fig. (11), we plot the random realizations of \( \Psi_x / \mathcal{S}_x \) as a function of \( D_{\text{PE}}[\mathcal{P}_\theta||\mathcal{P}_{\theta=0}] \) (circles) or \( \theta^2 \mathcal{I}(\theta = 0) \) (triangles), which are the denominators of HCRI and FRI, respectively. In Fig. (11), we plot a reciprocal function with the dashed line, which corresponds to the lower bound of HCRI (circles) or FRI (triangles). We see that all circles are located above the line, indicating that HCRI is satisfied for all of the realizations. On the other hand, some triangle points are below the line, which is a sign of the violation of FRI. This shows that for nonlinear systems, FRI can be violated in out of linear response regime.

**HCRI in stochastic oscillator.**— We next apply Eq. (6) to a stochastic limit cycle oscillator. Circadian clocks are biological limit cycle oscillators, prevalent in organisms, and they orchestrate activities of several organs. The temporal precision of circadian clocks is incredibly high (the standard deviation of the period is 3–5 min in 24 h) \([30]\), and several mechanisms have been proposed for this high precision \([31,34]\). At the same time, circadian clocks

\[ A_\theta(x, t) = \cdots \]

where \( \cdots \) is an arbitrary input function, and the product \( \circ \) should be interpreted in the Stratonovich sense.
and realizations are plotted by circles. Parameters are perturbed (orange line) dynamics, where the deterministic verification of HCRI in the stochastic oscillator. For randomly selected points, which are indicated by these studies, an explicit biochemical design principles for maximizing both of the sensitivity and the precision. Recently, Ref. [37] studied signals but also to noise, the precision and the sensitivity. A higher sensitivity are vulnerable not only to the periodic unc. 05 0 8 0 1 0 1 0.02 0.04 0.06 0.08 0.10 \( \phi(x) \) can be calculated by directly solving the ordinary differential equation [21]. The integrated phase from \( t = 0 \) to \( t = T \) is given by \( \int_0^T \phi(x(t)) dt \). Since the time derivative of \( \phi \) is \( \dot{\phi} = \sum_{i=1}^N (\partial_{x_i} \phi(x)) \circ \dot{x}_i \), we define a current \( \Theta(x) = \int_0^T \sum_{i=1}^N (\partial_{x_i} \phi(x)) \circ \dot{x}_i \) which yields the integrated phase from a trajectory \( \Gamma \). The temporal precision \( \mathcal{P}_{\phi} \) of the oscillator is defined as a function of \( \Theta(x) \), which is the variance of the phase. The sensitivity of the oscillator can be quantified by the phase difference between perturbed and unperturbed dynamics. Therefore we define the phase sensitivity \( \mathcal{S}_\phi \) as

\[
\mathcal{S}_\phi \equiv [\langle \phi(T) \rangle_{\theta} - \langle \phi(T) \rangle_{\theta=0}]^2.
\]

Figures 1(a) shows an illustration of the phase \( \phi(x) \) (the dotted line shows the isochron) and the phase difference between unperturbed and perturbed dynamics. From Eq. (6), \( \mathcal{P}_{\phi} \) and \( \mathcal{S}_\phi \) satisfy the following relation: \( \mathcal{P}_{\phi} / \mathcal{S}_\phi \geq 1 / \mathcal{D}_{\text{PE}}[\theta=0] \), which shows that when \( \mathcal{D}_{\text{PE}}[\theta=0] \) becomes larger, higher precision and higher sensitivity can be attained simultaneously. HCRI can be used to evaluate the efficiency, defined by \( \epsilon = \mathcal{D}_{\text{PE}}[\theta=0]^{-1} \mathcal{S}_\phi / \mathcal{P}_{\phi} \leq 1 \) which quantifies the performance of the perturbed oscillator.

We numerically confirm the inequality relation of HCRI. We consider the following two-dimensional noisy limit cycle oscillator:

\[
A_\theta(x, t) = \begin{bmatrix} x_1 - x_1^3 - x_2 + \theta u(t) \\ x_2 \end{bmatrix}, \quad B = \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}.
\]

where \( D \) is the noise intensity and \( u(t) = 0 \) for \( t > 0 \) and \( u(t) = 0 \) for \( t \leq 0 \). Using Monte Carlo simulations, we solve the Langvin equation of Eq. (14) with time resolution \( h = 0.0002 \) and evaluate the Pearson divergence \( \mathcal{D}_{\text{PE}}[\theta=0] \), the sensitivity \( \mathcal{S}_\phi \), and the precision \( \mathcal{P}_{\phi} \) (for details of the simulation, please see [21]). We randomly select \( D, \theta, \) and \( T \), and repeat simulations for \( N_S = 5.0 \times 10^6 \) times at each of the selected parameter settings (ranges of the random parameters are shown in the caption of Fig. 1(c)). For initial values, we randomly select a point on the closed orbit of the deterministic oscillation (the light blue line in Fig. 1(b)). Figure 1(c) plots \( \mathcal{P}_{\phi} / \mathcal{S}_\phi \) as a function of \( \mathcal{D}_{\text{PE}}[\theta=0] \), where points denotes random calculation of the ratio and the dashed line denotes \( 1 / \mathcal{D}_{\text{PE}}[\theta=0] \) corresponding to the saturating case of HCRI. As can be seen, all the points are located above the line, which empirically verifies HCRI.

Next, we show a relation between \( \mathcal{D}_{\text{PE}}[\theta=0] \) and the entropy production. Let \( \Delta S_m \) be the medium entropy have to synchronize to sunlight cycles so that the biological activities work at specific time. Since oscillators with higher sensitivity are vulnerable not only to the periodic signals but also to noise, the precision and the sensitivity are trade-off factors which is an uncertainty relation in stochastic oscillators. References [33, 36] showed that the phase-response curve of actual circadian clocks can be reproduced by simultaneous maximization of the sensitivity and the precision. Recently, Ref. [37] studied biochemical design principles for maximizing both of the quantities. Although the trade-off relation of the two quantities were indicated by these studies, an explicit inequality has not hitherto been reported.

We consider a deterministic limit cycle oscillator defined by \( \dot{x} = A(x) \). We can define the phase \( \phi \) on a closed orbit of the deterministic oscillation by \( \phi = \Omega \) where \( \Omega = 2\pi / \tau \) is the angular frequency of the oscillation (\( \tau \) is the period of the deterministic oscillation). In the presence of noise and an external signal, the dynamics obeys Eq. (11) with \( A_\theta(x, t) = A(x) + \theta u(t) \) where \( u(t) \in \mathbb{R}^N \) is the signal. Although \( \phi \) is defined only on the deterministic closed orbit, we can expand the definition of the phase onto the entire \( x \) space, which is denoted by \( \phi(x) \). We define the phase sensitivity as the period of the deterministic oscillation. (d) Medium entropy as a function of \( \theta \).
defined by $\Delta S_m = \left\langle \frac{1}{T} \sum_{i=1}^{N} \int_{0}^{T} A_i(x) \circ d \xi_i \right\rangle$ (when $T$ is sufficiently large, the boundary term can be ignored and $\Delta S_m \simeq \Delta S_{\text{tot}}$). Following the simulation procedure explained above, we calculate $\Delta S_m$ and $\Delta S_{\text{PE}} \{P_0 \Vert P_{\theta=0}\}$ (parameter settings are shown in the caption of Fig. 1(d)). In Fig. 1(d), we plot $\Delta S_m$ as a function of $\Delta S_{\text{PE}} \{P_0 \Vert P_{\theta=0}\}$ for fixed $\theta$ and $T$ [21]. We see that $\Delta S_m$ increases when $\Delta S_{\text{PE}} \{P_0 \Vert P_{\theta=0}\}$ increases, showing that larger Pearson divergence can be achieved for larger entropy production. When the stochastic oscillator can be approximated by the OU process around the deterministic orbit, Eq. (11) shows that the Pearson divergence increases exponentially when the noise intensity decreases. It is known that higher precision and higher sensitivity are achieved with higher entropy production, which is consistent with results above.

Conclusion.—In this Letter, we apply information inequalities to systems described by Langevin equations to obtain inequality relating the variance and the sensitivity in stochastic processes. We expect that this Letter bridges statistical inference theory, which has made remarkable progress in recent years, between stochastic thermodynamic systems, and provides a direction for applying the theory to obtain thermodynamic inequalities.

Acknowledgments.—This work was supported by MEXT KAKENHI Grant No. JP16K00325.

—hasegawa@biom.t.u-tokyo.ac.jp
—tan@biom.t.u-tokyo.ac.jp

[1] F. Ritort, in Advances in Chemical Physics, Vol. 137, edited by S. A. Rice (Wiley publications, 2008) pp. 31–123.
[2] U. Seifert, Rep. Prog. Phys. 75, 126001 (2012).
[3] C. Van den Broeck and M. Esposito, Physica A 418, 6 (2015).
[4] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa, Nat. Phys. 11, 131 (2015).
[5] A. C. Barato and U. Seifert, Phys. Rev. Lett. 114, 158101 (2015).
[6] A. C. Barato and U. Seifert, J. Phys. Chem. B 119, 6555 (2015).
[7] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Phys. Rev. Lett. 116, 120601 (2016).
[8] M. Polettini, A. Lazarescu, and M. Esposito, Phys. Rev. E 94, 052104 (2016).
[9] P. Pietzonka, A. C. Barato, and U. Seifert, Phys. Rev. E 93, 052145 (2016).
[10] J. M. Horowitz and T. R. Gingrich, Phys. Rev. E 96, 020103 (2017).
[11] K. Proesmans and C. V. den Broeck, EPL 119, 20001 (2017).
[12] P. Pietzonka, F. Ritort, and U. Seifert, Phys. Rev. E 96, 012101 (2017).
[13] S. Pigolotti, I. Neri, E. Roldán, and F. Jülicher, Phys. Rev. Lett. 119, 140604 (2017).
[14] A. Dechant and S.-i. Sasa, J. Stat. Mech: Theory Exp. 2018, 063209 (2018).
[15] A. Dechant and S.-i. Sasa, arXiv:1804.08250 (2018).
[16] H. Risken, Stochastic Processes: Methods of Solution and Applications, 2nd ed. (Springer, 1989).
[17] C. Gardiner, Stochastic methods: A Handbook for the Natural and Social Sciences (Springer, 2009).
[18] Y. Y. Chernyak, M. Chertkov, and C. Jarzynski, J. Stat. Mech: Theory Exp. 2006, P08001 (2006).
[19] P. C. Bressloff, in Stochastic Processes in Cell Biology, edited by P. C. Bressloff (Springer International Publishing, 2014) pp. 577–617.
[20] H. S. Wio, Path integrals for stochastic processes: An introduction (World Scientific, 2013).
[21] See Supplemental Material.
[22] Y. Tang, R. Yuan, and P. Ao, J. Chem. Phys. 141, 044125 (2014).
[23] A. B. Adib, J. Phys. Chem. B 112, 5910 (2008).
[24] A. Stuart, J. K. Ord, and S. Arnold, Classical Inference and the Linear Model, 6th ed., Kendall’s Advanced Theory of Statistics, Vol. 2A (Arnold, 1999).
[25] G. Casella and R. L. Berger, Statistical Inference, edited by C. Crockett (Duxbury, 2001).
[26] E. L. Lehmann and G. Casella, Theory of point estimation (Springer, 2003).
[27] R. Kawai, J. M. R. Parrondo, and C. V. den Broeck, Phys. Rev. Lett. 98, 080602 (2007).
[28] M. Esposito and C. Van den Broeck, EPL 95, 40004 (2011).
[29] M. Sugiyama, S. Liu, M. C. Du Plessis, M. Yamanaka, M. Yamada, T. Suzuki, and T. Kanamori, J. Comput. Sci. Eng. 7, 99 (2013).
[30] K. T. Moortgat, T. H. Bullock, and T. J. Sejnowski, J. Neurophysiol. 83, 971 (2000).
[31] A. T. Winfree, The Geometry of Biological Time, 2nd ed. (Springer, 2001).
[32] D. J. Needleman, P. H. E. Tiesinga, and T. J. Sejnowski, Physica D 155, 324 (2001).
[33] H. Kori, Y. Kawamura, and N. Masuda, J. Theor. Biol. 297, 61 (2012).
[34] Y. Hasegawa, Phys. Rev. E (2018), in press (arXiv:1712.09584).
[35] Y. Hasegawa and M. Arita, J. R. Soc. Interface 11, 20131018 (2014).
[36] Y. Hasegawa and M. Arita, Phys. Rev. Lett. 113, 108101 (2014).
[37] C. Fei, Y. Cao, Q. Ouyang, and Y. Tu, Nat. Commun. 9, 1434 (2018).
[38] Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence (Dover publications, Mineola, New York, 2003).
[39] T. Tomé, Braz. J. Phys. 36, 1285 (2006).
Supplemental Material for
“Information Inequality Interpretation of Uncertainty Relation in Stochastic Process”
Yoshihiko Hasegawa and Tan Van Vu

This supplementary material describes in detail the calculations introduced in the main text. Equation and figure numbers in this section are prefixed with S (e.g., Eq. (S1) or Fig. S1). Numbers without the prefix (e.g., Eq. (1) or Fig. 1) refer to items in the main text.

S1 Hammersley–Chapman–Robbins inequality (HCRI)

HCRI is a generalization of CRI, which was found after CRI [1, 2, 3]. We define the probability density functions $P_\theta(X)$ and $P_\vartheta(X)$, where $\theta, \vartheta \in \mathbb{R}$ are arbitrary parameters satisfying $\vartheta \neq \theta$. We notice

$$\langle P_\vartheta(X) P_\theta(X) - 1 \rangle_\theta = \int dX \left( P_\theta(X) - P_\vartheta(X) \right) = 0.$$  \hfill (S1)

From the property of Eq. (S1), the following relation holds:

$$\langle (\Theta(X) - \psi(\theta)) \left( \frac{P_\theta(X)}{P_\vartheta(X)} - 1 \right) \rangle_\vartheta = \langle (\Theta(X))_\vartheta - (\Theta(X))_\theta \rangle.$$  \hfill (S2)

Applying Cauchy–Schwarz inequality to Eq. (S2), we obtain HCRI:

$$\text{Var}_\theta [\Theta(X)] \geq \frac{\langle (\Theta(X))_\vartheta - (\Theta(X))_\theta \rangle^2}{\langle \left( \frac{P_\theta(X)}{P_\vartheta(X)} - 1 \right)^2 \rangle_\theta} = \frac{\langle (\Theta(X))_\vartheta - (\Theta(X))_\theta \rangle^2}{\mathbb{D}_{PE}[P_\theta||P_\vartheta]},$$  \hfill (S3)

where $\mathbb{D}_{PE}[P_\theta||P_\vartheta]$ is the Pearson divergence:

$$\mathbb{D}_{PE}[P_\theta||P_\vartheta] \equiv \int dX \left( \frac{P_\theta(X)}{P_\vartheta(X)} - 1 \right)^2 P_\vartheta(X) = \int dX \left( \frac{P_\vartheta(X)}{P_\theta(X)} \right)^2 P_\theta(X) - 1.$$  \hfill (S4)

Although HCRI is less tractable compared to CRI, the fundamental advantage of HCRI is that it does not require $P_\vartheta(X)$ to be differentiable and holds for arbitrary $\vartheta$. CRI is recovered in the limit of $\vartheta \to \theta + d\theta$. Applying a Taylor series expansion to $\mathbb{D}_{PE}[P_{\theta + d\theta}||P_{\theta}]$, we obtain

$$\mathbb{D}_{PE}[P_{\theta + d\theta}||P_{\theta}] = (d\theta)^2 \int \frac{\partial}{\partial \theta} \ln P_\theta(X)^2 P_\theta(X) dX = (d\theta)^2 I(\theta).$$  \hfill (S5)

Substituting Eq. (S5) into Eq. (S3), we prove CRI.
S2 Path integral

For readers’ convenience, we introduce the pre-point discretization procedure of the path integral after Refs. [4, 5]. Multivariate calculation is unnecessarily complicated; we here explain in a univariate case since multivariate extension is straightforward.

We consider the following Langevin equation (Ito interpretation)

$$\dot{x} = A_\theta(x, t) + \sqrt{2}C(x, t)\xi(t), \quad (S6)$$

where $\xi(t)$ is the white Gaussian noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = 2\delta(t - t')$. We discretize time by dividing the interval $[0, T]$ into $K$ equipartition intervals with time resolution $\Delta t$, where $T = K\Delta t$, $t_k = k\Delta t$, and $x^k \equiv x(t^k)$ (we use superscripts for specifying a point in temporal data). Discretization of Eq. (S6) yields

$$x^{k+1} - x^k = \Delta t A_\theta(x^k, t^k) + \sqrt{2}C(x^k, t^k)\Delta w^k, \quad (S7)$$

where $\Delta w^k$ is Wiener process with the following properties

$$\langle \Delta w^k \rangle = 0, \quad \langle \Delta w^k \Delta w^{k'} \rangle = \Delta t\delta_{kk'}.$$  

A stochastic trajectory $\mathcal{X} \equiv (x^1, x^2, ..., x^K)$ is specified given $\mathcal{W} \equiv (\Delta w^0, \Delta w^1, ..., \Delta w^{K-1})$ and $x^0$. The Wiener process $\Delta w^k$ has the following probability density function:

$$P(\mathcal{W}) = \prod_{k=0}^{K-1} P(\Delta w^k) = \prod_{k=0}^{K-1} \frac{1}{\sqrt{2\pi \Delta t}} \exp \left[ -\frac{(\Delta w^k)^2}{2\Delta t} \right]. \quad (S8)$$

Let us change variables of Eq. (S8) from $\mathcal{W} = (\Delta w^0, \Delta w^1, ..., \Delta w^{K-1})$ to $\mathcal{X} = (x^1, x^2, ..., x^K)$. From Eq. (S7), the determinant of a Jacobian matrix is

$$\left| \frac{\partial(x^1, ..., x^K)}{\partial(\Delta w^0, ..., \Delta w^{K-1})} \right| = \prod_{k=0}^{K-1} \sqrt{2B(x^k, t^k)}, \quad (S9)$$

where we used a fact that the determinant of triangular matrices is given by a product of their diagonal elements. Using Eqs. (S7), (S8), and (S9), we obtain

$$P_\theta(\mathcal{X}|x^0) = \left( \prod_{k=0}^{K-1} \frac{1}{\sqrt{4\pi \Delta t B(x^k, t^k)}} \right) \exp \left[ -\frac{1}{4}\sum_{k=0}^{K-1} \Delta t \left\{ \left( \frac{x^{k+1} - x^k}{\Delta t} - A_\theta(x^k, t^k) \right)^2 B(x^k, t^k) \right\} \right].$$

In the limit $K \to \infty$, $\mathcal{X} \to \Gamma \equiv [x(t)]_{t=0}^T$ and we write

$$P_\theta(\Gamma|x^0) = \mathcal{N} \exp \left[-\frac{1}{4} \int_0^T dt \left( \dot{x} - A_\theta(x, t) \right)^2 B(x, t)^{-1} \right], \quad (S10)$$

For an arbitrary function $g(x, t)$, the following relation holds

$$\int_0^T dt \left( \dot{x} - A_\theta(x, t) \right) g(x, t) = \left\{ \sum_{k=0}^{K-1} \Delta t \left\{ \left( \frac{x^{k+1} - x^k}{\Delta t} - A_\theta(x^k, t^k) \right) \right\} \right\} g(x^k, t^k) = \left\{ \sum_{k=0}^{K-1} \sqrt{2} \Delta w^k C(x^k, t^k) g(x^k, t^k) \right\} = 0. \quad (S11)$$

1 Instead of using the Jacobian, Ref. [5] used $\delta$ function to change the variables. Both of the approaches reduce the same result.
S3 Fisher information of trajectory density function

The log-likelihood is given by

\[
\ln \mathcal{P}_\theta(\Gamma) = \ln \mathcal{N} + \ln \mathcal{P}_\theta(x^0) - \frac{1}{4} \int_0^T dt \, (\dot{x} - A_\theta)^\top B^{-1} (\dot{x} - A_\theta).
\]  

(S12)

We calculate the second derivative of Eq. (S12):

\[
\frac{\partial^2}{\partial \theta^2} \ln \mathcal{P}_\theta(\Gamma) = \frac{\partial^2}{\partial \theta^2} \ln \mathcal{P}_\theta(x^0) - \frac{1}{2} \int_0^T dt \left( \frac{\partial}{\partial \theta} A_\theta \right)^\top B^{-1} \left( \frac{\partial}{\partial \theta} A_\theta \right) + \frac{1}{2} \int_0^T dt \, (\dot{x} - A_\theta)^\top B^{-1} \left( \frac{\partial^2}{\partial \theta^2} A_\theta \right).
\]

(S13)

When applying the expectation \( \langle \cdots \rangle_\theta \) to Eq. (S13), the last term vanishes due to Eq. (S11). Therefore, the Fisher information is given by Eq. (4) in the main text.

S4 Ito and Stratonovich currents

We show a relation between Ito and Stratonovich currents (cf. Eq. (S19)), both of which appear in the main text, for a univariate case (a multivariate generalization is shown below).

Ito and its equivalent Stratonovich Langevin equations are given by

\[
\begin{align*}
\dot{x} &= A(x) dt + \sqrt{2} C(x) \cdot dw, \tag{S14} \\
\dot{x} &= \left[ A(x) - C(x) C'(x) \right] dt + \sqrt{2} C(x) \circ dw, \tag{S15}
\end{align*}
\]

respectively, where \( w(t) \) is the Wiener process. In Eqs. (S14) and (S15), we use \( \cdot \) and \( \circ \) to explicitly express Ito and Stratonovich products, respectively. Let \( \beta(x) \) be an arbitrary function of \( x \). We are interested in a relation between the following two terms:

\[
U_S \equiv \int_0^T \beta(x) \circ dw, \quad U_I \equiv \int_0^T \beta(x) \cdot dw,
\]

where \( w(t) \) is the standard Wiener process. For \( K \to \infty \), their discretized representations are

\[
\begin{align*}
U_I &= \sum_{k=0}^{K-1} \beta(x^k) \Delta w^k, \tag{S16} \\
U_S &= \sum_{k=0}^{K-1} \beta\left(\frac{x^{k+1} + x^k}{2}\right) \Delta w^k, \tag{S17}
\end{align*}
\]

where \( \Delta w^k \equiv w(t^{k+1}) - w(t^k) \) is the Wiener process. Applying a Taylor series expansion and dropping terms whose order are higher than \( O(\Delta t) \), we obtain the following well-known relation [6]

\[
\begin{align*}
U_S &= \sum_{k=0}^{K-1} \left[ \beta(x^k) \Delta w^k + \frac{\sqrt{2}}{2} C(x^k) \beta'(x^k)(\Delta w^k)^2 \right] \\
&= \int_0^T \beta(x) \cdot dw + \frac{\sqrt{2}}{2} \int_0^T C(x) \beta'(x) \cdot dw^2 \\
&= U_I + \frac{\sqrt{2}}{2} \int_0^T C(x) \beta'(x) dt. \tag{S18}
\end{align*}
\]

where we used a relation \( dw^2 = dt \) in the last line, which is valid for any non-anticipating functions (see Chapter 4 in [6] for details).
Next we consider Ito and Stratonovich currents of the following forms:

\[ J_S \equiv \int_0^T \Lambda(x) \circ \dot{x} dt, \quad J_I \equiv \int_0^T \Lambda(x) \bullet \dot{x} dt, \]  

(S19)

where \( \Lambda(x) \) is an arbitrary projection function. Their discretized representations are

\[ J_I = \sum_{k=0}^{K-1} \Lambda(x^k) \left( x^{k+1} - x^k \right), \]  

(S20)

\[ J_S = \sum_{k=0}^{K-1} \Lambda(x^k) \left( \frac{x^{k+1} + x^k}{2} \right) \left( x^{k+1} - x^k \right), \]  

(S21)

Substituting Eqs. (S14) and (S15) into Eqs. (S20) and (S21), respectively, we obtain

\[ J_S = \int_0^T \Lambda(x) \left( A(x) - C(x)C'((x)) \right) dt + \sqrt{2} \int_0^T \Lambda(x)C(x) \circ dw, \]  

(S22)

\[ J_I = \int_0^T \Lambda(x)A(x) dt + \sqrt{2} \int_0^T \Lambda(x)C(x) \bullet dw. \]  

(S23)

Using Eq. (S18) \( \beta(x) = \Lambda(x)C(x) \), the following relation holds:

\[ \int_0^T \Lambda(x)C(x) \circ dw = \int_0^T \Lambda(x)C(x) \bullet dw + \sqrt{2} \int_0^T C(x) \frac{\partial \Lambda(x)C(x)}{\partial x} dt. \]  

(S24)

By substituting Eq. (S21) into Eq. (S22), a relation between the Stratonovich current \( J_S \) and the Ito current \( J_I \) is given by

\[ J_S = \int_0^T \Lambda(x) \left( A(x) - C(x)C'((x)) \right) dt + \sqrt{2} \left[ \int_0^T \Lambda(x)C(x) \bullet dw + \frac{\sqrt{2}}{2} \int_0^T C(x) \frac{\partial \Lambda(x)C(x)}{\partial x} dt \right] \]

\[ = \int_0^T \Lambda(x)A(x) dt + \sqrt{2} \int_0^T \Lambda(x)C(x) \bullet dw + \int_0^T \Lambda'(x)C(x)^2 dt \]

\[ = J_I + \int_0^T \Lambda'(x)C(x)^2 dt. \]  

(S25)

Therefore we find the following relation

\[ \int_0^T \Lambda(x) \circ \dot{x} dt = \int_0^T \Lambda(x) \bullet \dot{x} dt + \int_0^T B(x) \frac{\partial \Lambda(x)}{\partial x} dt. \]  

(S26)

For a multivariate case, we repeat the same calculations to obtain

\[ \int_0^T \sum_i \Lambda_i(x) \circ \dot{x}_i dt = \int_0^T \sum_i \Lambda_i(x) \bullet \dot{x}_i dt + \int_0^T \sum_{i,j} B_{ij}(x) \frac{\partial \Lambda_j(x)}{\partial x_i} dt. \]  

(S27)

Note that we confirmed Eq. (S27) for simple examples with Monte Carlo simulations.

**S5 Pre-point and mid-point discretizations**

In path integral, there exist subtleties regarding the discretization point. Typical schemes are pre-point, which is used in this Letter, and mid-point discretizations. Following Ref. [14], we here explain that both discretizations reduce to the same path integral representation for additive noise systems.

We consider a case that systems of interest has additive noise only, i.e., \( B(x,t) = B \). Although we explain for a univariate case, a multivariate generalization is straight-forward. Equation (S10) gives the pre-point
discretization representation, and Ref. [8] provided the the mid-point one. Both of the representations are given by

(midpoint) \( \mathcal{P}_\theta(\Gamma|x^0) = \mathcal{N} \exp \left[ \frac{1}{4} \int_0^T dt \left( \frac{\dot{x} - A_\theta(x,t)}{B} \right)^2 + \frac{1}{2} \frac{\partial}{\partial x} A_\theta(x,t) \right] \) (S28)

(prepoint) \( \mathcal{P}_\theta(\Gamma|x^0) = \mathcal{N} \exp \left[ -\frac{1}{4} \int_0^T dt \left( \frac{\dot{x} - A_\theta(x,t)}{B} \right)^2 \right] \) (S29)

Equations (S28) and (S29) seem to be inconsistent, since the midpoint representation includes an additional term. However, in Eqs. (S28) and (S29), \( \int (\dot{x} - A_\theta(x,t))^2 dt \) has a cross term \( \int A_\theta(x,t) \dot{x} dt \), which should be interpreted differently for mid-point and pre-point discretizations [7]:

\[
\int_0^T A_\theta(x,t) \dot{x} dt = \left\{ \begin{array}{ll}
\int_0^T A_\theta(x,t) \bullet \dot{x} dt & \text{(prepoint)}, \\
\int_0^T A_\theta(x,t) \circ \dot{x} dt & \text{(midpoint)}. 
\end{array} \right.
\]

Using Eq. (S25), we obtain (replace \( \Lambda(x) \) with \( A_\theta(x,t) \))

\[
\int A_\theta(x,t) \circ \dot{x} dt = \int A_\theta(x,t) \bullet \dot{x} dt + \int_0^T B \frac{\partial}{\partial x} A_\theta(x,t) dt. \tag{S30}
\]

Substituting Eq. (S30) into Eq. (S28), Eq. (S28) agrees with Eq. (S29). Therefore, both of the discretizations reduce to the same path-integral representation (this also holds for a multivariate case).

For systems with multiplicative noise [i.e., \( B(x,t) \) depends on \( x \)], the situation becomes complicated. As for the path integral representations for the multiplicative case, different studies proposed different representations (for instance, Refs. [8, 9] employed slightly different expressions, both of which reduce to the identical expression for the additive noise case). This indicates that there seems no consensus for the multiplicative path integral representation.

**S6 Equality condition of TUR**

As explained in the main text, the projection \( \Lambda(x) \) which satisfies the equality condition of TUR does not exist for arbitrary \( A(x) \) and \( B(x) \). Still, there is a possibility that a system of interest can achieve the equality condition of TUR by properly determining \( A(x) \), \( B(x) \), and \( \Lambda(x) \), simultaneously. Here we consider the condition of \( \Lambda(x) \), \( A(x) \) and \( B(x) \) to achieve the equality condition of TUR in one dimensional case using CRI.

The drift vector of the modified dynamics is

\[
A_\theta(x) \equiv (\theta + 1)A(x) - \frac{\theta}{P^{ss}(x)} \frac{\partial}{\partial x} B(x) P^{ss}(x).
\]

The average of generalized current of the original dynamics is

\[
\dot{j} \equiv \left\langle \int_0^T dt \Lambda(x) \circ \dot{x} \right\rangle = TJ^{ss} \int dx \Lambda(x), \tag{S31}
\]

where \( J^{ss} \) is constant at steady-state in one-dimensional case. When \( J^{ss} = 0 \), the entropy production vanish, i.e., \( j = \Delta S_{tot} = 0 \). Therefore, we assume here that \( J^{ss} \neq 0 \) and demand the system to have periodic boundary conditions. Since the steady-state distribution remains unchanged, the current of the modified dynamics is

\[
J_\theta^{ss} = \left[(1 + \theta)A(x) - \frac{\theta}{P^{ss}(x)} \partial_x B(x) P^{ss}(x)\right] P^{ss}(x) - \partial_x B(x) P^{ss}(x) = (1 + \theta)J^{ss},
\]
which yields $\psi(\theta) = (1 + \theta)J$. Converting from Ito to Stratonovich-type currents by using Eq. (S26), we have

$$
\partial_0 \ln \mathcal{P}_0(\Gamma) \propto \int_0^T dt \left[ J^{ss} B(x)^{-1} \cdot \dot{x} - J^{ss} B(x)^{-1} A(x) \right]
$$

$$
= \int_0^T dt \left[ J^{ss} B(x)^{-1} \circ \dot{x} - B(x) \frac{\partial}{\partial x} \left( J^{ss} B(x)^{-1} \right) - J^{ss} B(x)^{-1} A(x) \right]
$$

$$
= \int_0^T dt J^{ss} B(x)^{-1} \circ \dot{x} - \theta \int_0^T dt \left[ J^{ss} B(x)^{-1} \right]
$$

$$
- \int_0^T dt \left[ B(x) \frac{\partial}{\partial x} \left( J^{ss} B(x)^{-1} \right) + J^{ss} B(x)^{-1} A(x) \right].
$$

(S32)

From the equality condition of CRI [cf. Eq. (5)], Eq. (S32) should be expressed as

$$
\partial_0 \ln \mathcal{P}_0(\Gamma) \propto \int_0^T dt \Lambda(x) \circ \dot{x} - \psi(\theta)
$$

$$
= \int_0^T dt \Lambda(x) \circ \dot{x} - (1 + \theta)J.
$$

(S33)

Since correspondence between Eqs. (S32) and (S33) should hold for arbitrary trajectory $\Gamma$ to attain the equality condition, the following relations must be satisfied:

$$
\Lambda(x) = \mu(\theta) \frac{J^{ss} B(x)^{-1}}{P^{ss}(x)},
$$

(S34)

$$
j = \mu(\theta) \int_0^T dt \frac{J^{ss} B(x)^{-1}}{P^{ss}(x)^2},
$$

(S35)

$$
\dot{j} = \mu(\theta) \int_0^T dt \left[ B(x) \frac{\partial}{\partial x} \left( J^{ss} B(x)^{-1} \right) + J^{ss} B(x)^{-1} A(x) \right],
$$

(S36)

where $\mu(\theta)$ is a scaling function of $\theta$, and $j \equiv \left< \int_0^T dt \Lambda(x) \circ \dot{x} \right> = \mu(\theta) T [J^{ss} \int \frac{B(x)^{-1}}{P^{ss}(x)} dx]$. Equations (S35) and (S36) hold for arbitrary trajectories if and only if

$$
\frac{[J^{ss} B(x)^{-1}]}{P^{ss}(x)^2} = \frac{j}{\mu(\theta)},
$$

(S37)

$$
B(x) \frac{\partial}{\partial x} \left( J^{ss} B(x)^{-1} \right) + J^{ss} B(x)^{-1} A(x) = \frac{j}{\mu(\theta)},
$$

(S38)

It can be observed that from Eq. (S37), Eq. (S38) is calculated into

$$
B(x) \frac{\partial}{\partial x} \left( J^{ss} B(x)^{-1} \right) + J^{ss} B(x)^{-1} A(x) = \frac{j}{\mu(\theta)} J^{ss} \left[ A(x) P^{ss}(x) + B(x) \frac{\partial}{\partial x} P^{ss}(x) \right]
$$

$$
= \frac{j}{\mu(\theta)} J^{ss} \left[ A(x) P^{ss}(x) - \frac{\partial}{\partial x} B(x) P^{ss}(x) \right] = \frac{j}{\mu(\theta)},
$$

where, from the first to the second line, we used $\partial_x B(x) P^{ss}(x)^2 = P^{ss}(x) \left[ 2 B(x) \partial_x P^{ss}(x) + P^{ss}(x) \partial_x B(x) \right] = 0$ [cf. Eq. (S37)]. From Eq. (S37), we get $P^{ss}(x) = c/\sqrt{B(x)}$, where $c > 0$ is a normalizing constant. From $J^{ss} = A(x) P^{ss}(x) - \partial_x B(x) P^{ss}(x)$, we obtain $A(x)$ as follows:

$$
A(x) = \frac{J^{ss} + \partial_x B(x) P^{ss}(x)}{P^{ss}(x)} = \kappa_1 \sqrt{B(x)} + \frac{1}{2} \frac{\partial}{\partial x} B(x),
$$

(S39)

where $\kappa_1$ is a parameter. In summary, the equality is satisfied if and only if $A(x)$ and $\Lambda(x)$ are given by

$$
A(x) = \kappa_1 \sqrt{B(x)} + \partial_x B(x)/2,
$$

(S40)

$$
\Lambda(x) = \kappa_2 / \sqrt{B(x)},
$$

(S41)
Figure S1: The random realizations of $\frac{2(\Theta(\Gamma))^2}{\Delta S_{\text{tot}} \cdot \text{Var}[\Theta(\Gamma)]}$ are plotted as a function of $\text{Var}[^2]$. (a) The random realizations are plotted for Cases A (circles) and B (triangles), which satisfy Eqs. (S40) and (S41). (b) The random realizations are plotted for Cases C (circles), D (triangles), and E (squares), which do not satisfy Eqs. (S40) and (S41).

which are Eqs. (8) and (9) in the main text.

We confirm the condition of Eqs. (S40) and (S41) by Monte Carlo simulations. We test the following five cases:

- (Case A) $B(x) = 2 + \sin(x)$, $A(x) = \kappa_1 \sqrt{B(x)} + \frac{1}{2} \frac{dB(x)}{dx}$, $\Lambda(x) = \frac{1}{\sqrt{B(x)}}$.
- (Case B) $B(x) = 2 + \sin(x) + \sin(2x)$, $A(x) = \kappa_1 \sqrt{B(x)} + \frac{1}{2} \frac{dB(x)}{dx}$, $\Lambda(x) = \frac{1}{\sqrt{B(x)}}$.
- (Case C) $B(x) = 2 + \sin(x) + \sin(2x)$, $A(x) = \kappa_1 \sqrt{B(x)} + \frac{1}{2} \frac{dB(x)}{dx}$, $\Lambda(x) = 1$.
- (Case D) $B(x) = 2 + \sin(x) + \sin(2x)$, $A(x) = \kappa_1 \sqrt{B(x)}$, $\Lambda(x) = \frac{1}{\sqrt{B(x)}}$.
- (Case E) $B(x) = 2 + \sin(x) + \sin(2x)$, $A(x) = \kappa_1$, $\Lambda(x) = \frac{1}{\sqrt{B(x)}}$.

Cases A and B satisfy Eqs. (S40) and (S41), whereas Cases C, D, and E do not. We randomly select $\kappa_1 \in [0.5, 5.0]$ and $T \in [0.5, 2.0]$, and numerically solve Langevin equations with time resolution $h = 0.001$. We repeat $N_S = 1.0 \times 10^6$ times for each of the selected parameter settings to calculate the current $\Theta(\Gamma) = \int_0^T \Lambda(x) \circ \dot{x} dt$ and the total entropy production $\Delta S_{\text{tot}}$. Figure S1(a) plots $\frac{2(\Theta(\Gamma))^2}{\Delta S_{\text{tot}} \cdot \text{Var}[\Theta(\Gamma)]}$ as a function of $\text{Var}[\Theta(\Gamma)]$ for Cases A (circles) and B (triangles), and (b) shows those for Cases C (circles), D (triangles), and E (squares). $\frac{2(\Theta(\Gamma))^2}{\Delta S_{\text{tot}} \cdot \text{Var}[\Theta(\Gamma)]}$ should be no larger than 1 according to TUR, and TUR saturates when the value is 1. The random realizations of Cases A and B are very close to 1, and this shows that Cases A and B are saturating cases of TUR as predicted by Eqs. (S40) and (S41). On the other hand, Cases C, D, and E, which do not satisfy the equality condition of Eqs. (S40) and (S41), show the values smaller than 1, demonstrating that these cases are not saturating cases. These results validate the correctness of Eqs. (S40) and (S41).

S7 Bounds for Ornstein–Uhlenbeck process

The Pearson divergence is calculated analytically for Ornstein–Uhlenbeck (OU) process. In the main text, we have considered the following OU process

$$\dot{x} = -\alpha x + \theta u(t) + \sqrt{2D} \xi(t).$$

(S42)
Therefore, the Pearson divergence is given by

$$\Delta x^k = [-\alpha x^k + \theta u^k] \Delta t + \sqrt{2D} \Delta w^k.$$  

The probability of the discretized trajectory $X = [x^1, x^2, ..., x^K]$ is given by

$$P_0(X|x^0) = \frac{1}{(4\pi D\Delta t)^{K/2}} \exp \left[ -\frac{1}{4D\Delta t} \sum_{k=0}^{K-1} (x^{k+1} - x^k - (-\alpha x^k + \theta u^k) \Delta t)^2 \right].$$

The Pearson divergence between $P_0(X)$ and $P_{\theta=0}(X)$ is

$$\mathbb{D}_{PE}[P_0||P_{\theta=0}] = \int \prod_{k=0}^{K} dx^k \left[ \frac{P_0(X|x^0)P_0(x^0)}{P_{\theta=0}(X|x^0)P_{\theta=0}(x^0)} \right]^2 P_{\theta=0}(X|x^0)P_{\theta=0}(x^0) - 1.$$ 

Let us introduce new variables $y^k (k = 1, 2, ..., K)$ defined by

$$y^{k+1} = x^{k+1} - x^k + \alpha x^k \Delta t.$$  

The determinant of a Jacobian is

$$\left| \frac{\partial (y^1, y^2, ..., y^K)}{\partial (x^1, x^2, ..., x^K)} \right| = 1.$$  

Using Eqs. (S43) and (S44), the probability density of $Y \equiv [y^1, y^2, ..., y^K]$ is

$$P_0(Y|x^0) = \frac{1}{(4\pi D\Delta t)^{K/2}} \exp \left[ -\frac{1}{4D\Delta t} \sum_{k=0}^{K-1} (y^{k+1} - \theta u^k \Delta t)^2 \right].$$

Therefore, the Pearson divergence is given by

$$\mathbb{D}_{PE}[P_0||P_{\theta=0}] = \int dx^0 \int \prod_{k=1}^{K} dy^k \left[ \frac{P_0(Y|x^0)P_0(x^0)}{P_{\theta=0}(Y|x^0)P_{\theta=0}(x^0)} \right]^2 P_{\theta=0}(Y|x^0)P_{\theta=0}(x^0) - 1$$

$$= -1 + \int dx^0 \left( \frac{P_0(x^0)}{P_{\theta=0}(x^0)} \right)^2 P_{\theta=0}(x^0) \int \prod_{k=1}^{K} \frac{dy^k}{(4\pi D\Delta t)^{K/2}}$$

$$\times \exp \left[ -\frac{1}{2D\Delta t} \sum_{k=0}^{K-1} (y^{k+1} - \theta u^k \Delta t)^2 + \frac{1}{2D\Delta t} \sum_{k=0}^{K-1} (y^k - 1)^2 - \frac{1}{4D\Delta t} \sum_{k=0}^{K-1} (y^{k+1} - 1)^2 \right].$$

$$= -1 + \exp \left[ \sum_{k=0}^{K-1} \left( \frac{\theta^2}{2D} \Delta t \right)^2 \right] \int dx^0 \left( \frac{P_0(x^0)}{P_{\theta=0}(x^0)} \right)^2 P_{\theta=0}(x^0).$$

When the initial distributions are the same for $\theta \neq 0$ and $\theta = 0$, in the limit of $K \to \infty$, we obtain

$$\mathbb{D}_{PE}[P_0||P_{\theta=0}] = -1 + \exp \left[ \frac{\theta^2}{2D} \int_0^T u(t)^2 dt \right]. \quad (S45)$$

Next, we prove inequality relations of Eq. (12). Using $\exp(x) \geq 1 + x$ in Eq. (S45), we can easily show

$$\mathbb{D}_{PE}[P_0||P_{\theta=0}] \geq \theta^2 T(\theta = 0) = \frac{\theta^2}{2D} \int_0^T u(t)^2 dt. \quad (S46)$$

When $x = 0$ at time $t = 0$, the variance and the mean of Eq. (S42) is given by

$$\langle x(T) \rangle_\theta = \left[ \theta \int_0^T u(s)e^{\alpha s} ds \right] e^{-\alpha T},$$

$$\text{Var}_{\theta=0}[x(T)] = \frac{D}{\alpha} \left[ 1 - e^{-2\alpha T} \right].$$
For any $u(t) \geq 0$, using Cauchy–Schwarz inequality, we have
\[
\left[ \int_0^T u(s)e^{\alpha s} ds \right]^2 \leq \int_0^T u(s)^2 ds \int_0^T e^{2\alpha s} ds
= \frac{e^{2\alpha T} - 1}{2\alpha} \int_0^T u(s)^2 ds,
\]  
which yields the following relation:
\[
\frac{\mathcal{Q}_x}{\mathcal{E}_x} \geq \frac{1}{\theta^2 I(\theta = 0)}.
\]  
Equations (S46) and (S48) prove Eq. (12) in the main text.

\section*{S8 Numerical simulation details}

In the main text, we performed numerical simulations to confirm the inequality bound. We explain its implementations in this section.

\subsection*{S8.1 Monte Carlo simulations}

We carried out Monte Carlo simulations for the stochastic limit cycle. We used Eq. (S7) to solve Ito Langevin equations (this method is known as the Euler–Maruyama scheme). Stratonovich-type currents are calculated by Eq. (S21).

We numerically calculated the Pearson divergence. We generate trajectories from the Langevin equations with parameter $\theta$. Let $N_S$ be the number of generated trajectories and $\mathcal{X}_i$ be the $i$th realization of the trajectories. Then the integral of the Pearson divergence is approximated by the following summation:
\[
\mathbb{D}_{PE} [P_\theta||P_0] \approx \frac{1}{N_S} \sum_{i=1}^{N_S} \left( \frac{P_\theta(\mathcal{X}_i)}{P_0(\mathcal{X}_i)} - 1 \right)^2,
\]  
where
\[
P_\theta(\mathcal{X}|x_0) = \exp \left[ -\frac{\Delta t}{4} \sum_{k=0}^{K-1} \sum_{i,j} \left( \frac{x_{i,k+1} - x_{i,k}}{\Delta t} - A_{\theta,i} (x^{k},t^k) \right) B^{-1}_{ij}(x^{k},t^k) \left( \frac{x_{j,k+1} - x_{j,k}}{\Delta t} - A_{\theta,j} (x^{k},t^k) \right) \right].
\]  
Here $B^{-1}_{ij}(x^{k},t^k)$ is an $i,j$th element of $B(x^{k},t^k)^{-1}$ and we omitted $\mathcal{N}$ because it cancels out in Eq. (S49).

\subsection*{S8.2 Definition of phase}

We can define the phase for limit cycle oscillators \cite{10}. For deterministic oscillators, we can define the phase $\phi$ on a closed orbit by
\[
\frac{d\phi}{dt} = \Omega,
\]  
where $\Omega$ is the angular frequency of the deterministic oscillation. We can expand the definition of the phase into an entire space $x \in \mathbb{R}^N$. Let $x_a$ be a point on the closed orbit and $x_b$ be a point that is not on the orbit. According to Eq. (S50), we can determine $\phi(x_a)$. Since the closed orbit is attractor in limit cycle oscillators, $x_b$ eventually converges to the closed orbit for $t \to \infty$. We let $x_a$ and $x_b$ time-evolve for the same time duration. If the two points eventually converge to the same point on the closed orbit, then we can assign $\phi(x_b) = \phi(x_a)$. 

9
References

[1] A. Stuart, J. K. Ord, and S. Arnold. *Classical Inference and the Linear Model*, volume 2A of *Kendall’s Advanced Theory of Statistics*. Arnold, 6 edition, 1999.

[2] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, 2001.

[3] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer, 2003.

[4] H. S. Wio. *Path integrals for stochastic processes: An introduction*. World Scientific, 2013.

[5] P. C. Bressloff. The WKB method, path-integrals, and large deviations. In P. C. Bressloff, editor, *Stochastic Processes in Cell Biology*, pages 577–617. Springer International Publishing, 2014.

[6] C. Gardiner. *Stochastic methods: A Handbook for the Natural and Social Sciences*. Springer, 2009.

[7] A. B. Adib. Stochastic actions for diffusive dynamics: Reweighting, sampling, and minimization. *J. Phys. Chem. B*, 112:5910–5916, 2008.

[8] V. Y. Chernyak, M. Chertkov, and C. Jarzynski. Path-integral analysis of fluctuation theorems for general Langevin processes. *J. Stat. Mech: Theory Exp.*, 2006:P08001, 2006.

[9] Y. Tang, R. Yuan, and P. Ao. Summing over trajectories of stochastic dynamics with multiplicative noise. *J. Chem. Phys.*, 141:044125, 2014.

[10] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover publications, Mineola, New York, 2003.