Abstract. Uncertainty quantification is a primary challenge for reliable modeling and simulation of complex stochastic dynamics. Precisely due to the complex nature of such problems they are typically plagued with incomplete information that may enter as uncertainty in the model parameters, or even in the model itself. Furthermore, due to their dynamic nature, we need to assess the impact of these uncertainties in the transient and long-time behavior of the stochastic models and derive corresponding uncertainty bounds for observables of interest. A special class of such challenges is parametric uncertainties in the model and in particular sensitivity analysis along with the corresponding sensitivity bounds for the stochastic dynamics. Finally, sensitivity analysis can be further complicated in models with a high number of parameters that render straightforward approaches, such as gradient methods, impractical.

In this paper, we derive uncertainty and sensitivity bounds for path-space observables of stochastic dynamics in terms of suitable information theoretic divergences such as relative entropy rate and path Fisher Information Matrix. These bounds are tight, depend on the variance of the particular observable and are computable typically through single-model Monte Carlo simulation. In the case of sensitivity analysis, the sensitivity bounds rely on the path Fisher Information Matrix, hence they are also gradient-free and allow for computationally efficient sensitivity bounds, even in systems with a high number of parameters.

Version: March 18, 2015

Key words. relative entropy, relative entropy rate, path-space Fisher information matrix, uncertainty quantification, sensitivity analysis

AMS subject classifications. 65C05

1. Introduction. In this paper, we derive uncertainty and sensitivity bounds for path-space observables of stochastic dynamics in terms of suitable information theoretic divergences such as relative entropy rate and path Fisher Information Matrix. Reliable modeling and simulation of complex systems often suffers from incomplete information that may enter as uncertainty in the model parameters, or even in the model itself. We develop an approach that provides uncertainty bounds for observables of interest in the transient and long-time behavior of the stochastic models. The presented method also yields bounds on parametric sensitivity for the stochastic dynamics. It is particularly useful in realistic stochastic models, for example, biochemical reaction networks, which are characterized by a high number of parameters that render straightforward sensitivity analysis approaches, such as gradient methods, impractical. We present sensitivity bounds that are computable and sufficiently sharp.

A main novelty of the presented results is their application to cases where the model is represented by a path measure for a Markov process \( \{X_t\}_{t \geq 0} \). Thus the proposed method allows us to compute bounds also for the weak error of path-dependent quantities.

Estimating sensitivity indices appears as a common task in many applications ranging from engineering and financial mathematics to biochemistry. Standard methods that apply Monte Carlo simulations to estimate directly [1, 3] include finite-difference approximations combined with coupling methods [1, 2], likelihood ratio methods [9], polynomial chaos expansions, etc. The bounds we present avoid expensive Monte Carlo simulations of the sensitivity indices by providing error bounds for them. The derived bounds are based on Fisher information and are obtained from different inequalities and representations of relative entropy (or Kullback-Leibler divergence). It is also useful to provide bounds based on Fisher information because Fisher information is extensively utilized in optimal experimental design, as well as in statistics, for estimation, identifiability, etc. Moreover, in order to obtain the tightest possible bounds, it is crucial to find the optimal constant that multiplies the Fisher information in these inequalities.

*This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Number DE-SC-0010539 (P.D.), DE-SC-0010723 (M.A.K, Y.P.), DE-SC-0010549 (P.P.).

1Division of Applied Mathematics, Brown University, Providence, RI 02912, USA, dupuis@dam.brown.edu

2Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305, USA, markos@math.umass.edu

3Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305, USA, yannis@math.umass.edu

4Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, USA, plechac@math.udel.edu
These results rely in part on an upper bound derived recently in [6] and a companion lower bound in [16], for functionals of probability measures $P \in \mathcal{P}(\Omega)$ and $Q \in \mathcal{P}(\Omega)$, where $Q$ is viewed as the “true” probabilistic model, and $P$ is a computationally tractable “nominal” or “reference” model. In this paper we start our analysis by showing that these inequalities, for bounded observables $f$ of random variables with probabilities $P$ and $Q$, are easily rewritten in the form
\[
\Xi_-(Q, P; f) \leq \mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \Xi_+(Q, P; f),
\]
where $\Xi_+(Q, P; f) \geq 0$, $(\Xi_-(Q, P; f) \leq 0)$ and $\Xi_+(Q, P; f) = 0$ if and only if $P = Q$ or $f$ is deterministic a.s. with respect to $Q$. Furthermore, $\Xi_+(Q, P; f)$ depend on the relative entropy (Kullback-Leibler divergence) between $P$ and $Q$. These bounds are of general nature and characterize the errors incurred if one uses the more computationally tractable $\mathbb{E}_P[f]$ instead of $\mathbb{E}_Q[f]$. The bounds on sensitivity indices for parametric model families $P^\theta$ then follow by asymptotic expansions of $Q \equiv P^{\theta+\nu}$ in $\nu$, which is a straightforward procedure under smoothness assumptions when the parameter is finite dimensional, i.e., $\theta \in \mathbb{R}^k$.

As discussed in Section 2 bounds of the type (1.1) can be derived from different divergences used to discriminate between two probability measures $P$ and $Q$. For example, a first natural choice is based on the Csiszar-Kullback-Pinsker (CKP) inequality, which bounds the total variation norm by the relative entropy $\mathcal{R}(Q || P)$, see (2.27). Another approach uses $\chi^2$-divergence (or Pearson divergence) and derives a bound by a direct application of Cauchy-Schwarz inequality, see (2.28).

The bounds (1.1) presented in this paper are based on the variational characterization used in [6]. The variational approach guarantees optimal constants in the estimates and thus tighter bounds in terms of the relative entropy than those obtained in (2.27) or (2.28).

In the context of parametrized models the general bounds (1.1) give a tool for estimating sensitivity of observables to perturbations in model parameters. More precisely, given a parametric family of probability measures $P^\theta(\omega)$, $\theta \in \mathbb{R}^k$, on the common measurable space $(\Omega, \mathcal{B})$, we study bounds on perturbations of $\mathbb{E}_{P^\theta}[f]$ under changes of $P^\theta$. By the term “weak error” we mean errors in averages or expected values for various classes of functions. The derived sensitivity bounds can be viewed as sharp and computable bounds on stability constants for the weak errors for bounded continuous functions $f \in C_b(\Omega)$, in cases when the measure $P^\theta$ is approximated by $P^{\theta+\nu}$, under assumptions of smoothness on the mapping $\theta \mapsto P^\theta$. The mapping defines a finite dimensional submanifold parametrized by $\theta \in \mathbb{R}^k$ of the manifold of probability measures $\mathcal{P}(\Omega)$ on $\Omega$. We establish estimates of the type
\[
|\mathbb{E}_{P^{\theta+\nu}}[f] - \mathbb{E}_{P^\theta}[f]| \leq C_f \Phi(P^\theta)(\phi(\nu)),
\]
where the constant $C_f$ depends only on the function $f$, the dependence on $\nu$ is isolated in the function $\phi$ and $\Phi(\cdot)$ is a functional depending on the measure $P^\theta$.

If $P(\omega)$ is smooth and $k = 1$ then the perturbations of observables are characterized by
\[
|\mathbb{E}_{P^{\theta+\nu}}[f] - \mathbb{E}_{P^\theta}[f]| = S_f(P)\epsilon + o(\epsilon),
\]
where we define the sensitivity index
\[
S_f(P) = \frac{d}{d\theta} \mathbb{E}_{P^\theta}[f].
\]

The sensitivity indices in the case of $\theta \in \mathbb{R}^k$ are defined in the natural way in terms of $\partial_\theta \mathbb{E}_{P^\theta}[f]$.

With stochastic dynamics in mind, we consider a stochastic process $\{X_t\}_{t \geq 0}$ with the stationary measure $\mu(dx)$, and a process $\{Y_t\}_{t \geq 0}$ with the initial measure $\nu(dx)$, and we denote by $P = P_{[0,T]} Q = Q_{[0,T]}$ the respective measures on the path space. We consider as an observable a measurable functional $F(X)$ of the process $\{X_t\}_{0 \leq t \leq T}$. An observable of great interest is the ergodic average
\[
F(X) = \frac{1}{T} \int_0^T f(X_s) \, ds
\]
for a test function $f$. Under proper ergodic assumptions one has
\[
\lim_{T \to \infty} F(X) = \mathbb{E}_\mu[f] \equiv \int f \, d\mu.
\]
The derived bounds are now set on path space and characterize the errors incurred when approximating $\mathbb{E}_{Q[0,T]}[\mathcal{F}]$ by $\mathbb{E}_{P[0,T]}[\mathcal{F}]$:

$$\Xi_-(Q[0,T]; P[0,T]; \mathcal{F}) \leq \mathbb{E}_{Q[0,T]}[\mathcal{F}] - \mathbb{E}_{P[0,T]}[\mathcal{F}] \leq \Xi_+(Q[0,T]; P[0,T]; \mathcal{F}).$$

(1.5)

where again $\Xi_\pm(Q[0,T]; P[0,T]; \mathcal{F}) \geq 0$, (resp. $\Xi_- (Q[0,T]; P[0,T]; \mathcal{F}) \leq 0$) and $\Xi_\pm(Q[0,T]; P[0,T]; \mathcal{F}) = 0$ if and only if $P[0,T] = Q[0,T]$ or $\mathcal{F}$ is deterministic a.s. Furthermore, $\Xi_\pm(Q[0,T]; P[0,T]; \mathcal{F})$ are given by a variational formula that depends on the path relative entropy between $Q[0,T]$ and $P[0,T]$,

$$\mathcal{R} (Q[0,T] \mid P[0,T]) = \mathbb{E}_{Q[0,T]} \left[ \log \frac{dQ[0,T]}{dP[0,T]} \right].$$

We demonstrate that the relative entropy on the path space, at least in the stationary regime and for interesting models, is an easily computable quantity, implying in turn that the bounds in (1.5) are computable using Monte Carlo simulation based only on the model $P$.

The bound (1.5) can be linearized to

$$|\mathbb{E}_{Q[0,T]}[\mathcal{F}] - \mathbb{E}_{P[0,T]}[\mathcal{F}]| \leq \sqrt{\frac{1}{T} \text{Var}_{P[0,T]}[\mathcal{T} \mathcal{F}]} + \sqrt{\frac{2}{T} \mathcal{R} (Q[0,T] \mid P[0,T]) + \mathcal{O} \left( \frac{1}{T} \mathcal{R} (Q[0,T] \mid P[0,T]) \right)},$$

(1.6)

Note that the relative entropy and variance scale linearly with $T$. An implication of the linearized bound is that it allows us to provide sensitivity analysis bounds which are general and valid in both transient and long-time regimes.

2. Uncertainty quantification information inequalities and sensitivity bounds.

2.1. Distances and divergences of probability measures. Bounds of the type (1.5) are based on characterizing a distance or divergence between the measures, $P^\theta$, $P^{\theta + \epsilon}$, under which the averages are evaluated. While our primary goal is to characterize the bounds based on relative entropy, other divergences can be also used to derive similar bounds with different level of sharpness.

**Definition 2.1.** The total variation norm between two probability measures $Q$ and $P$ on $(\Omega, \mathcal{B})$ is defined by

$$\|Q - P\|_{TV} = \sup_{A \in \mathcal{B}} |Q(A) - P(A)|.$$  

(2.1)

We consider also two pseudo-distances, or divergences in the statistics terminology.

**Definition 2.2.** For two probability measures $Q$, $P$ on $(\Omega, \mathcal{B})$ the relative entropy (information divergence, Kullback-Leibler divergence) of $Q$ from $P$ is defined by

$$\mathcal{R} (Q \mid P) = \left\{ \begin{array}{ll} \int \log \frac{dQ}{dP}(\omega) Q(d\omega) = \int \frac{dQ}{dP}(\omega) \log \frac{dQ}{dP}(\omega) P(d\omega), & \text{if } Q \ll P \text{ and } \frac{dQ}{dP} \log \frac{dQ}{dP} \text{ is } P\text{-integrable}, \\ +\infty & \text{otherwise}. \end{array} \right.$$  

(2.2)

The Kullback-Leibler divergence is a particular case of a family of Csiszár $\phi$-divergences which are functionals of the form

$$\mathcal{R}_\phi (Q \mid P) = \left\{ \begin{array}{ll} \int \phi \left( \frac{dQ}{dP}(\omega) \right) P(d\omega), & \text{if } Q \ll P \text{ and } \phi \left( \frac{dQ}{dP} \right) \text{ is } P\text{-integrable}, \\ +\infty & \text{otherwise}. \end{array} \right.$$  

(2.3)

with a convex function $\phi : \mathbb{R}^+ \to \mathbb{R}$, $\phi(1) = \phi(0)$. In the case of the relative entropy we have $\phi(x) = x \log x$. Another choice of the convex function, $\phi(x) = (x - 1)^2$, gives a member of the $\phi$-divergence family known as $\chi^2$-divergence.

**Definition 2.3.** The $\chi^2$ divergence of two probability measures $Q$, $P$ on $(\Omega, \mathcal{B})$ is defined by

$$\chi^2 (Q \mid P) = \left\{ \begin{array}{ll} \left( \frac{dQ}{dP}(\omega) - 1 \right)^2 P(d\omega), & \text{if } Q \ll P, \\ +\infty & \text{otherwise}. \end{array} \right.$$  

(2.4)
2.2. Information inequalities and observable-based divergence. We turn to a variational formulation that provides sharp estimates in terms of relative entropy. Let $\mathcal{M}(\Omega)$ denote the measurable functions from $\Omega$ into $\mathbb{R}$ and let $\mathcal{M}_b(\Omega)$ be the subset of functions that are uniformly bounded. For $f \in \mathcal{M}_b(\Omega)$ and $c \in \mathbb{R}$ we introduce the cumulative moment generating function (logarithmic generating function)

$$\Lambda_{P,f}(c) = \log \mathbb{E}_P[e^{cf}] = \log \int e^{cf} \, dP. \quad (2.5)$$

We restrict our analysis to the functions $f$ for which $\Lambda_{P,f}(c)$ is finite at least in a neighborhood of the origin. More specifically, we have the following definition of the set $\mathcal{E}$.

**Definition 2.4.** A function $f \in \mathcal{M}_b(\Omega)$ belongs to the set $\mathcal{E}$ if and only if there exists $c_0 > 0$ such that $\Lambda_{P,f}(\pm c_0) < \infty$.

The properties of $\Lambda_{P,f}$ then guarantee that $\Lambda_{P,f}(c)$ is finite for all $c \in [-c_0, c_0]$. We note that $\mathbb{E}_P[|f|]$ is finite for all $f \in \mathcal{E}$. It will be more convenient to work with the cumulative generating function of the centered observable $\bar{f} \equiv f - \mathbb{E}_P[f]$

$$\bar{\Lambda}_{P,f}(c) = \log \mathbb{E}_P[e^{c(f-\mathbb{E}_P[f])}] = \log \int e^{c(f-\mathbb{E}_P[f])} \, dP. \quad (2.6)$$

Recalling the basic properties of the cumulant generating function for $f \in \mathcal{E}$ that is not essentially constant, we have that $\bar{\Lambda}_{P,f}(\cdot)$ is a strictly convex function which is $C^\infty$ in a neighborhood of the origin, with the derivatives $\bar{\Lambda}_{P,f}^{(k)}(0)$ defining the cumulants of $f - \mathbb{E}_P[f]$ under $P$. In particular, $\bar{\Lambda}_{P,f}(0) = \bar{\Lambda}'_{P,f}(0) = 0$ and $\bar{\Lambda}''_{P,f}(0) = \text{Var}[f]$. The following characterization of exponential integrals is well-known in statistics and large deviation theory (see e.g., [7]).

**Lemma 2.5.** Let $f \in \mathcal{M}_b(\Omega)$ and $P$ be a probability measure on $(\Omega, \mathcal{B})$. Then

$$\log \mathbb{E}_P[e^f] = \sup_{Q \ll P} \{ \mathbb{E}_Q[f] - \mathcal{R}(Q \| P) \}. \quad (2.7)$$

**Proof.** It suffices to consider only $Q$ such that $\mathcal{R}(Q \| P) < \infty$ in (2.7). Let the probability measure $R$ be defined by $dR/dP = e^f/\mathbb{E}_P[e^f]$. If $\mathcal{R}(Q \| P) < \infty$, then $Q \ll P$ implies $Q \ll R$. Thus

$$-\mathcal{R}(Q \| P) + \mathbb{E}_Q[f] = -\mathbb{E}_Q \left[ \log \left( \frac{dQ}{dP} \right) \right] + \mathbb{E}_Q[f]$$

$$= -\mathbb{E}_Q \left[ \log \left( \frac{dQ}{dR} \right) \right] - \mathbb{E}_Q \left[ \log \left( \frac{dR}{dP} \right) \right] + \mathbb{E}_Q[f]$$

$$= -\mathcal{R}(Q \| R) + \log \mathbb{E}_P[e^f].$$

Now use that $\mathcal{R}(Q \| R) \geq 0$ and $\mathcal{R}(Q \| R) = 0$ if and only if $Q = R$ [7 Lemma 1.4.1]. This establishes (2.7) and also shows that $R$ is the supremizing measure. \[ \square \]

By changing $f$ to $c(f - \mathbb{E}_P[f])$, we obtain a variational formula for the cumulant generating function:

$$\bar{\Lambda}_{P,f}(c) = \sup_{Q \ll P} \{ c(\mathbb{E}_Q[f] - \mathbb{E}_P[f]) - \mathcal{R}(Q \| P) \}. \quad (2.8)$$

The variational characterization gives us the following upper and lower bounds for $f \in \mathcal{M}_b(\Omega)$ and $c > 0$:

$$\mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \frac{1}{c} \log \mathbb{E}_P[e^{c(f-\mathbb{E}_P[f])}] + \frac{1}{c} \mathcal{R}(Q \| P), \quad (2.9)$$

$$\mathbb{E}_Q[f] - \mathbb{E}_P[f] \geq -\frac{1}{c} \log \mathbb{E}_P[e^{-c(f-\mathbb{E}_P[f])}] - \frac{1}{c} \mathcal{R}(Q \| P). \quad (2.10)$$

These inequalities can be extended to any $f \in \mathcal{E}$, and we give the argument for the case of the upper bound [2.9]. Recall that $f \in \mathcal{E}$ implies $\mathbb{E}_P[|f|] < \infty$. If $\mathbb{E}_P[e^{c(f-\mathbb{E}_P[f])}] = \infty$, then (2.9) holds automatically. If $\mathbb{E}_P[e^{c(f-\mathbb{E}_P[f])}] < \infty$, let $f^{a,b} = [f \vee (-a)] \wedge b$ for $a,b \in \mathbb{R}$, and apply (2.9) with $f - \mathbb{E}_P[f]$ replaced by
\( f_{a,b} - \mathbb{E}_P[f] \). First let \( a \to \infty \) and use the Monotone Convergence Theorem, and then send \( b \to \infty \) and use the dominating function \( e^{c(f - \mathbb{E}_P[f])} \) to obtain (2.9) as written.

Using these inequalities we obtain tight estimates as in Chowdhary and Dupuis, 3, and Li and Xie, 10 by optimizing over \( c > 0 \):

\[
\sup_{c>0} \left\{ -\frac{1}{c} \tilde{\Lambda}_{P,f}(-c) - \frac{1}{c} \mathcal{R}(Q \| P) \right\} \leq \mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \inf_{c>0} \left\{ \frac{1}{c} \tilde{\Lambda}_{P,f}(c) + \frac{1}{c} \mathcal{R}(Q \| P) \right\}.
\]  
\[(2.11)\]

We refer to upper and lower bounds of this form as Uncertainty Quantification Information Inequalities. The corresponding bounds define a new type of divergence between probability measures \( P \) and \( Q \) as well as the observable \( f \). More precisely, based on (2.11) we give the following definitions.

**Definition 2.6.** For any two probability measures \( P \) and \( Q \) with \( \mathcal{R}(Q \| P) < \infty \) and any observable \( f \in \mathcal{E} \), we define

\[
\Xi_+(Q, P; f) = \inf_{c>0} \left\{ \frac{1}{c} \tilde{\Lambda}_{P,f}(c) + \frac{1}{c} \mathcal{R}(Q \| P) \right\},
\]  
\[(2.12)\]

and similarly

\[
\Xi_-(Q, P; f) = \sup_{c>0} \left\{ -\frac{1}{c} \tilde{\Lambda}_{P,f}(-c) - \frac{1}{c} \mathcal{R}(Q \| P) \right\}.
\]  
\[(2.13)\]

Then the bounds (2.11) are rewritten as

\[
\Xi_-(Q, P; f) \leq \mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq \Xi_+(Q, P; f),
\]  
\[(2.14)\]

We show that \( \Xi_{\pm}(Q, P; f) \) has the properties of a divergence, such as the relative entropy \( \mathcal{R}(Q \| P) \) and the \( \chi^2 \)-divergence (2.4). However, it also captures the role of fluctuations of the observable \( f \), as is further quantified in Theorem 2.11 below.

**Theorem 2.7 (Representation).** If \( f \in \mathcal{E} \) and \( \mathcal{R}(Q \| P) < \infty \) then we have

\[
\Xi_+(Q, P; f) = \tilde{\Lambda}'_{P,f}(F^{-1}(\mathcal{R}(Q \| P))) \quad \text{and} \quad \Xi_-(Q, P; f) = \tilde{\Lambda}_{P,f}(-F^{-1}(\mathcal{R}(Q \| P))),
\]  
\[(2.15)\]

where

\[ F(c) := -\tilde{\Lambda}_{P,f}(c) + c \tilde{\Lambda}'_{P,f}(c) \]

is a strictly increasing function on \((0, \bar{c})\), and where \( \bar{c} = \sup\{c : \tilde{\Lambda}_{P,f}(c) < \infty\} \).

**Proof.** Let \( \Theta_+(c; \rho) \equiv \frac{1}{c} \tilde{\Lambda}_{P,f}(c) + \frac{1}{c} \rho^2 \), where \( \rho^2 = \mathcal{R}(Q \| P) \). Then

\[
\Xi_+(P, Q; f) = \inf_{c>0} \Theta_+(c; \rho)
\]  
\[(2.16)\]

If \( \rho \neq 0 \) then since \( \tilde{\Lambda}_{P,f}(0) = \tilde{\Lambda}'_{P,f}(0) = 0 \) and \( \tilde{\Lambda}_{P,f} \) is strictly convex, \( \Theta_+(c; \rho) \) tends to \( \infty \) as \( c \downarrow 0 \) and as \( c \uparrow \infty \). Hence the infimum is achieved. Suppose an infimum of \( A > 0 \) is achieved at \( 0 < c_1 < c_2 < \infty \), so that \( \tilde{\Lambda}_{P,f}(c_i) + \rho^2 = c_i A, \ i = 1,2 \). If \( \bar{c} = (c_1 + c_2) / 2 \), then the strict convexity of \( \tilde{\Lambda}_{P,f} \) implies \( \tilde{\Lambda}_{P,f}(\bar{c}) + \rho^2 < c A \). This contradicts the minimality of \( c_i \), and thus shows the minimizer is unique. Since \( \tilde{\Lambda}_{P,f}(0) = \tilde{\Lambda}'_{P,f}(0) = 0 \) we can continuously extend the function \( \Theta_+(c,0) \) to \( c = 0 \) by \( \Theta_+(0,0) = 0 \). Then by direct calculation and lower semicontinuity the problem \( (P_\nu) \) extended to \( c \geq 0 \) has the unique minimizer \( c^* = 0 \) with the minimum value equal to 0. Since \( \tilde{\Lambda}_{P,f}(\cdot) \) is a proper convex function and \( C^\infty \) in its domain of finiteness we have, for all \( \rho \in \mathbb{R} \), the optimality condition

\[
-\frac{1}{c^2} \tilde{\Lambda}_{P,f}(c) + \frac{1}{c} \tilde{\Lambda}'_{P,f}(c) - \frac{1}{c^2} \rho^2 = 0.
\]  
\[(2.17)\]

Multiplying (2.17) by \( c^2 \), we obtain that the minimizer \( c^* = c^*(\rho) \) satisfies

\[
- \tilde{\Lambda}_{P,f}(c) + c \tilde{\Lambda}'_{P,f}(c) = \rho^2.
\]  
\[(2.18)\]
Due to the strict convexity of $\tilde{\Lambda}_{P,f}(c)$, we have that $F'(c) = c\tilde{\Lambda}'_{P,f}(c) > 0$ for $c > 0$, thus $F = F(c)$ is strictly increasing and invertible. Therefore, from (2.18) we have:

$$c^* = c^*(\rho) = F^{-1}(\rho^2). \quad (2.19)$$

Substituting in (2.16) and using (2.18), we have that

$$\Xi_+(Q, P; f) = \Theta_+(c^*(\rho); \rho) = \tilde{\Lambda}'_{P,f}(c^*(\rho)) = \tilde{\Lambda}'_{P,f}(F^{-1}(\rho^2)). \quad (2.20)$$

The representation of the lower bound $\Xi_-(Q, P; f) = \Lambda'_{P,f}(F^{-1}(\rho^2))$ is computed in a similar way. □

From the proof above we deduce that the dependence on the cumulant generating function of $\rho$ can be removed if a bound is available. Note that if $\Psi : \mathbb{R} \to \mathbb{R}$ is convex with a minimum at zero at the origin, then in the definition of $\Psi^*(t)$, its Legendre-Fenchel transform, the supremum can be restricted to $(0, \infty)$.

**Corollary 2.8.** Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a convex and continuously differentiable function such that $\Psi(0) = \Psi'(0) = 0$ and

$$\Lambda_{P,f}(c) \equiv \log \mathbb{E}_P[e^{c(f - \mathbb{E}_P[f])}] \leq \mathbb{E}_P[e^{c(f - \mathbb{E}_P[f])}] = \Lambda_{P,f}(c).$$

and define $\Psi^*_+(t) = (\Psi^*_+)^{-1}(t)$ as the (generalized) inverse of the Legendre-Fenchel transform $\Psi^*(t) = \sup_{c > 0}\{ct - \Psi(c)\}$ of the function $\Psi$. Then

$$\Xi_Q[f] - \Xi_P[f] \leq \Psi^*_+(\mathcal{R}(Q || P)). \quad (2.21)$$

The quantities $\Xi_\pm$ have properties similar to other types of divergence between two probability measures and also characterize dependence on the observable $f$.

**Proposition 2.9 (Observable-based divergence).** If $f \in \mathcal{E}$, then

(i) $\Xi_+(Q, P; f) \geq 0$ and $\Xi_-(P, Q; f) \leq 0$.

(ii) $\Xi_+(Q, P; f) = 0$ if and only if $Q = P$ or $f$ is constant $P$-a.s.

**Proof.** The proofs for $\Xi_+$ and $\Xi_-$ are similar and therefore we prove only the former case.

(i) The proof uses the fact that both terms in the variational definition of $\Xi_+$,

$$\Xi_+(Q, P; f) = \inf_{c > 0}\left\{ \frac{1}{c} \Lambda_{P,f}(c) + \frac{1}{c} \mathcal{R}(Q || P) \right\},$$

are non-negative. The relative entropy $\mathcal{R}(Q || P)$ is a divergence hence always non-negative, and thus $\frac{1}{c} \mathcal{R}(Q || P) \geq 0$ for all $c > 0$. Furthermore, by Jensen’s inequality,

$$\frac{1}{c} \Lambda_{P,f}(c) \equiv \frac{1}{c} \log \mathbb{E}_P\left[e^{c(f - \mathbb{E}_P[f])}\right] \geq \frac{1}{c} \log e^{\mathbb{E}_P[f]} = \mathbb{E}_P[f - \mathbb{E}_P[f]] = 0.$$

(ii) If $f = \mathbb{E}_P[f]$ then $\Lambda_{P,f}(c) \equiv 0$. Since $\mathcal{R}(Q || P) \in [0, \infty)$,

$$\Xi_+(Q, P; f) = \inf_{c > 0}\left\{ \frac{1}{c} \mathcal{R}(Q || P) \right\} = 0.$$

If $Q = P$ then $\mathcal{R}(Q || P) = 0$ and

$$0 \leq \Xi_+(Q, P; f) = \inf_{c > 0}\left\{ \frac{1}{c} \Lambda_{P,f}(c) \right\} \leq \lim_{c \to 0} \frac{1}{c} \Lambda_{P,f}(c) = \Lambda_{P,f}(0) = 0.$$

If $\Xi_+(Q, P; f) = 0$ then, as in the proof of Theorem 2.7, we have that there exists $c^* \geq 0$ such that

$$\Xi_+(Q, P; f) = \frac{1}{c^*} \Lambda_{P,f}(c^*) + \frac{1}{c^*} \mathcal{R}(Q || P) = 0,$$

where both terms are non-negative. Thus $\mathcal{R}(P || Q) = 0$, implying that $Q = P$ and $\Lambda_{P,f}(c^*) = 0$. However, when $\mathcal{R}(P || Q) = 0$ then $c^* = 0$ and $\Lambda_{P,f}(0) = 0$ for any $f$. On the other hand, if $c^* > 0$ then $\mathcal{R}(P || Q) \neq 0$.
and to have $\Xi_+ = 0$ it is required that $c^* \to +\infty$. Thus $\lim_{c \to \infty} \bar{\Lambda}_P(f)(c) = 0$ and $\bar{\Lambda}_P(f)(0) = 0$, and together with $\Lambda_P(f)(c) \geq 0$ and strict convexity this implies that $\Lambda_P(f)(c) \equiv 0$. In other words

$$\log \mathbb{E}_P \left[ e^{c(f - \mathbb{E}_P[f])} \right] = 0, \text{ for all } c \geq 0.$$  

For $\epsilon > 0$ define the set $A_\epsilon = \{ \omega \mid f - \mathbb{E}_P[f] > \epsilon \}$ and assume $P(A_\epsilon) > 0$. Then we have for any $c > 0$

$$\log \int e^{c(f - \mathbb{E}_P[f])} dP > \log \int_{A_\epsilon} e^c dP = ce + \log P(A_\epsilon),$$

and choosing $\tilde{c} = -\log P(A_\epsilon)/\epsilon$ we have the contradiction $\log \int e^{\tilde{c}(f - \mathbb{E}_P[f])} dP > 0$. Thus $f - \mathbb{E}_P[f] = 0$ $P$-a.e. and $\Lambda_{P,f}(c) \equiv 0$ which means that $\inf_{c > 0} \frac{1}{c} \mathcal{R}(Q \| P) = 0$ even if $\mathcal{R}(Q \| P) \neq 0$. \[ \square \]

2.3. Linearization of the UQ bounds. The UQ bounds (2.13) and the representations (2.15) can be made more explicit in terms of the asymptotic expansion at $\mathcal{R}(Q \| P) = 0$, i.e. when $Q$ is a perturbation $P$. We first prove an asymptotic expansion for the solution of the optimization problems in (2.11).

**Lemma 2.10.** For two probability measures $P, Q$ on $(\Omega, \mathcal{B})$ set $\rho^2 = \mathcal{R}(Q \| P)$. Assume $\rho^2 < \infty$ and $f \in \mathcal{E}$. Then there exists a function $c^*(\rho)$ which is the unique solution of

$$(P_+) \quad \inf_{c > 0} \left\{ \frac{1}{c} \Lambda_{P,f}(c) + \frac{1}{c} \mathcal{R}(Q \| P) \right\}$$

as well as

$$(P_-) \quad \sup_{c > 0} \left\{ -\frac{1}{c} \bar{\Lambda}_{P,f}(-c) - \frac{1}{c} \mathcal{R}(Q \| P) \right\}.$$  

Furthermore, there is $\rho_0 > 0$ such that the optimal solution $c^*(\rho)$ is $C^\infty$ in $(0, \rho_0)$ and admits the expansion

$$c^*(\rho) = c^*_1 \rho + O(\rho^2), \quad (2.22)$$

where

$$c^*_1 = \sqrt{\frac{2}{\text{Var}_P[f]}}. \quad (2.23)$$

**Proof.** We first solve $(P_+)$. Let $\Theta_+(c; \rho) \equiv \frac{1}{c} \Lambda_{P,f}(c) + \frac{1}{c} \rho^2$. Following Theorem 2.7, we obtain the optimality condition (2.17). Multiplying (2.17) by $c$, we define

$$G(c, \rho) := -\frac{1}{c} \bar{\Lambda}_{P,f}(c) + \bar{\Lambda}_{P,f}(c) - \frac{1}{c} \rho^2. \quad (2.24)$$

Next, we apply the Implicit Function Theorem at $c = 0, \rho = 0$ as follows; first we have that

$$\frac{\partial}{\partial c} G(c, 0) = \frac{1}{2} \bar{\Lambda}''_{P,f}(0) + O(c),$$

and thus obtain

$$\lim_{c \to 0} \frac{\partial}{\partial c} G(c, 0) = \frac{1}{2} \bar{\Lambda}''_{P,f}(0) = \text{Var}_P[f].$$

Since $\text{Var}_P[f] > 0$, by the Implicit Function Theorem there exists a unique solution $c^*(\rho) > 0$, $c^*(0) = 0$ of $G(c, \rho) = 0$ (and thus of (2.17)) and $c^*(\rho) \in C^\infty$ for $\rho$ in a neighborhood of the origin. Differentiating $G(c^*(\rho), \rho) = 0$ and setting $\rho = 0$ yields terms in the Taylor expansion of $c^*(\rho)$. In particular, using the notation $\dot{c}^* = dc^*/d\rho$, we have $c^*(\rho) \bar{\Lambda}''_{P,f}(c^*(\rho)) \dot{c}^*(\rho) = 2\rho$, and thus by setting $\dot{c}^*(0) = \lim_{\rho \to 0^+} \dot{c}^*(\rho)$ we have $(\dot{c}^*(0))^2 = 2/\bar{\Lambda}''_{P,f}(0)$, which concludes the proof by observing again that $\text{Var}_P[f] = \bar{\Lambda}''_{P,f}(0)$. 

To prove that \( c^*(\rho) \) is also the solution of \( (P_-) \) we observe that

\[
\sup_{c>0} \left\{ -\frac{1}{c} \hat{\Lambda}_{P,f}(-c) - \frac{1}{c} \mathcal{R}(Q \| P) \right\} = -\inf_{c>0} \left\{ \frac{1}{c} \hat{\Lambda}_{P,f}(-c) + \frac{1}{c} \mathcal{R}(Q \| P) \right\},
\]

and using the same arguments as for \( (P_+) \) we conclude that the unique solution is obtained as the solution of the optimality condition

\[-\frac{1}{c^2} \hat{\Lambda}_{P,f}(-c) - \frac{1}{c} \hat{\Lambda}'_{P,f}(-c) - \frac{1}{c^2} \rho^2 = 0, \quad c > 0,
\]

which is, under the change of the variable \( c \to -c \), the same as (2.17) and thus analogous calculations yield the result.

Next, substituting the expansion in \( \rho \) for the optimal value (2.22) we obtain asymptotics in \( \rho^2 = \mathcal{R}(Q \| P) \) of the upper and lower bounds for the UQ error (2.14).

**Theorem 2.11 (Linearization).** Under the assumption \( f \in \mathcal{E} \), we have

(i) the asymptotic expansion \( \Xi_\pm(P, Q; f) = \pm \sqrt{\text{Var}_P[f]} \sqrt{2\mathcal{R}(Q \| P)} + \mathcal{O}(\mathcal{R}(Q \| P)) \),

(ii) an estimate of the weak error

\[
|\mathbb{E}_Q[f] - \mathbb{E}_P[f]| \leq \sqrt{\text{Var}_P[f]} \sqrt{2\mathcal{R}(Q \| P)} + \mathcal{O}(\mathcal{R}(Q \| P)).
\]

If needed, the term \( \mathcal{O}(\mathcal{R}(Q \| P)) \) can be further resolved using the asymptotic expansions of \( c^*(\rho) \) and \( \Theta_{\pm}(c; \rho) \) defined in Lemma 2.10 in terms of \( \rho^2 = \mathcal{R}(Q \| P) \).

**Proof.** The proof follows from the Taylor expansion of (2.15) in \( \rho \), where \( \rho^2 = \mathcal{R}(Q \| P) \), around \( \rho = 0 \). First, we note that \( \hat{\Lambda}_{P,f}(0) = \hat{\Lambda}'_{P,f}(0) = 0 \) and \( \hat{\Lambda}''_{P,f} = \text{Var}_P[f] \). Therefore \( F^{-1}(0) = 0 \), and the upper bound becomes

\[
\Xi_+(Q, P; f) = \hat{\Lambda}'_{P,f}(F^{-1}(\rho^2)) = \hat{\Lambda}'_{P,f}(0) + \hat{\Lambda}''_{P,f}(0) F^{-1}(\rho^2) + \mathcal{O}(|F^{-1}(\rho^2)|^2).
\]

We conclude using (2.14) and the expansion (2.22).

### 2.4. UQ bounds for other pseudo-distances

We can use other (pseudo)distances between measures and obtain bounds of the type (1.2) with different constants \( C_f \) and functionals \( \Phi \). In particular, we can apply the inequality between the relative entropy and \( \chi^2 \) divergence. If \( P, Q \) are two probability measures on \((\Omega, \mathcal{B})\) and \( Q \ll P \), then \( \mathcal{R}(Q \| P) \leq \chi^2(Q \| P) \). Furthermore, the Csiszár-Kullback-Pinsker inequality states

\[
\|Q - P\|_{TV} \leq \sqrt{2\mathcal{R}(Q \| P)}.
\]

Detailed discussion and proofs can be found, e.g., in [20]. Using \( \|Q - P\|_{TV} = \sup_{\|f\|_\infty \leq 1} \{\mathbb{E}_Q[f] - \mathbb{E}_P[f]\} \) and the Csiszár-Kullback-Pinsker inequality (2.26) we obtain

\[
|\mathbb{E}_Q[f] - \mathbb{E}_P[f]| \leq \|f\|_{\infty} \sqrt{2\mathcal{R}(Q \| P)}.
\]

This bound is of the type (1.2), however, its application to the perturbation case with the measures \( P^\theta \) and \( P^{\theta+v} \) may not lead to the best estimate due to the potentially large constant \( C_f = \|f\|_{\infty} \). The constant \( C_f \) can be improved by using the \( \chi^2 \)-divergence for bounding the weak error.

**Theorem 2.12.** Let \( P, Q \) be two probability measures on \((\Omega, \mathcal{B})\) with \( Q \ll P \). If \( f \in \mathcal{M}_b(\Omega) \) then

\[
|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq \sqrt{\text{Var}_P[f]} \sqrt{\chi^2(Q \| P)}.
\]
The leading term in this expansion is a quadratic form defined by the Fisher Information Matrix $\mathbf{I}$. Assume also that suitable dominating functions for various functions of $\omega$ in a neighborhood of the origin. Then for $(i)$ Then $c^* \in (0, \infty)$ such that $\int p^\theta \log \frac{p^{\theta + \epsilon \omega}}{p^\theta} \, d\theta = \log \frac{p^{\theta + \epsilon \omega}}{p^\theta}$. Let $p^\theta(\omega) = \frac{dQ^\theta(\omega)}{dP^\theta}$. Then there is a measurable set $N \subset \Omega$ such that $R(N) = 0$, and such that for all $\omega \not\in N$ the mapping $\theta \mapsto p^\theta(\omega)$ from $\mathbb{R}^k$ to $(0, \infty)$ is $C^3$. Where needed, we also assume the existence of suitable dominating functions for various functions of $p^\theta$.

Under Condition 2.1 the relative entropy can be expressed as

$$\mathcal{R}(p^{\theta + \omega} \| p^\theta) = \int p^{\theta + \omega}(\omega) \log \frac{p^{\theta + \omega}(\omega)}{p^\theta(\omega)} \, R(d\omega).$$

Using the Taylor expansion and the fact $\int [\partial_\theta \log p^\theta(\omega)] p^\theta(\omega) \, R(d\omega) = 0$, we have the perturbative expansion

$$\mathcal{R}(p^{\theta + \omega} \| p^\theta) = \frac{1}{2} \sum_{ij} v_i v_j \int \frac{1}{p^\theta(\omega)} [\partial_{\theta_i} p^\theta(\omega)][\partial_{\theta_j} p^\theta(\omega)] R(d\omega) + \mathcal{O}(|v|^3). \quad (2.29)$$

The leading term in this expansion is a quadratic form defined by the Fisher Information Matrix $\mathbf{I}$(2.28, 2.29) $\equiv \int [\partial_{\theta_i} p^\theta(\omega)][\partial_{\theta_j} p^\theta(\omega)] R(d\omega) = - \int [\partial_{\theta_i} \log p^\theta(\omega)] p^\theta(\omega) R(d\omega). \quad (2.30)$

We apply the derived bounds of Theorem 2.11 for the weak error in order to obtain bounds on the Sensitivity Indices

$$S_{f,v}(p^\theta) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} (\mathbb{E}_{p^{\theta + \epsilon \omega}}[f] - \mathbb{E}_{p^\theta}[f]). \quad (2.31)$$

Lemma 2.13. Assume Condition 2.1 and let $v \in \mathbb{R}^k$.

(i) Then

$$\mathcal{R}(p^{\theta + v} \| p^\theta) = \frac{1}{2} \sum_{ij} \mathbf{I}(p^\theta)_{ij} v_i v_j + \mathcal{O}(|v|^3). \quad (2.32)$$

(ii) Assume also that $f \in \mathcal{E}$ and thus the cumulant generating function $\tilde{\Lambda}_{p^\theta,f}(c) \equiv \log \mathbb{E}_{p^\theta}[e^{c(f - \mathbb{E}_{p^\theta}f)}]$ exists in a neighborhood of the origin. Then for $v \in \mathbb{R}^k$ and $c$ in a neighborhood of the origin there exists a function $c^*(c)$ which is the unique solution of

$$\mathbb{P} \inf_{c > 0} \left\{ \frac{1}{c} \tilde{\Lambda}_{p^\theta,f}(c) + \frac{1}{c} \mathcal{R}(p^{\theta + \epsilon v} \| p^\theta) \right\}, \quad (P_+)$$

as well as

$$\mathbb{P} \sup_{c > 0} \left\{ -\frac{1}{c} \tilde{\Lambda}_{p^\theta,f}(-c) - \frac{1}{c} \mathcal{R}(p^{\theta + \epsilon v} \| p^\theta) \right\}. \quad (P_-)$$
Furthermore the function $c^*(\epsilon)$ admits the perturbation expansion

$$
c^*(\epsilon) = c_1^* \epsilon + O(\epsilon^2),
$$

where

$$
c_1^* = \sqrt{\frac{\sum_{ij} I(P^\theta)_{ij} v_i v_j}{\text{Var}_{p^\theta}[f]}}.
$$

Proof. The claim in (i) follows from (2.28) and (2.29). The claim in (ii) follows directly from Lemma 2.10 after expanding the relative entropy in $\epsilon$, i.e., writing $\rho^\theta(\epsilon) = \epsilon^2 \frac{1}{2} \sum_{ij} I(P^\theta)_{ij} v_i v_j + O(\epsilon^3)$. Substituting in (2.22) and (2.23) we obtain (2.33) and (2.34). □

As a direct consequence of Theorem 2.11 we obtain a bound on the sensitivity indices by substituting $c^*(\epsilon)$ from (2.33) into $\Theta_{\pm}(c, \rho)$ (see the proof of Lemma 2.10).

**Theorem 2.14.** Under the assumptions of Lemma 2.13, it holds that for $v \in \mathbb{R}^k$ and $\epsilon \neq 0$

$$
\frac{1}{|\epsilon|} \left| \mathbb{E}_{P^{\theta+\epsilon v}}[f] - \mathbb{E}_{P^\theta}[f] \right| \leq \sqrt{\text{Var}_{p^\theta}[f] \sum_{ij} I(P^\theta)_{ij} v_i v_j + O(\epsilon)},
$$

or equivalently

$$
|S_{f,v}(P^\theta)| \leq \sqrt{\text{Var}_{p^\theta}[f] \sum_{ij} I(P^\theta)_{ij} v_i v_j}.
$$

The bound (2.36) relates likelihood-ratio sensitivity analysis methods, [9], that develop efficient estimators for the sensitivity indices (2.31), with information-theory based methods, showing that the latter provide a sensitivity bound on (2.31). We refer to the inequality (2.36) as a Sensitivity Bound.

**Remark 2.1.** The bound (2.36) on the sensitivity index is a direct consequence of more general non-infinitesimal bounds such as Theorem 2.11. We note that in the special case of sensitivity analysis, where we consider small perturbations in the parameter space, we can obtain sensitivity bounds of the same form as (2.36), directly from the Cauchy-Schwarz inequality:

$$
|S_{f,v}(P^\theta)| = \left| \frac{d}{d\epsilon} \mathbb{E}_{P^{\theta+\epsilon v}}[f] \right|
$$

$$
= \left| \int (f(\omega) - \mathbb{E}_{P^\theta}[f]) \left( \frac{d}{d\epsilon} \log p^{\theta+\epsilon v}(\omega) \right) p^{\theta}(\omega) R(d\omega) \right|
$$

$$
\leq \left( \int (f(\omega) - \mathbb{E}_{P^\theta}[f])^2 p^{\theta}(\omega) R(d\omega) \right)^{1/2} \left( \int \left( \frac{d}{d\epsilon} \log p^{\theta+\epsilon v}(\omega) \right)^2 p^{\theta}(\omega) R(d\omega) \right)^{1/2}
$$

$$
= \sqrt{\text{Var}_{p^\theta}[f]} \sqrt{\sum_{ij} I(P^\theta)_{ij} v_i v_j}.
$$

Finally, we can also use (2.28) applied to $P^\theta$ and $P^{\theta+\epsilon v}$ and obtain the same bound as in (2.36).

**Remark 2.2.** The Cramer-Rao inequality (2.30), [12], is a special case of the sensitivity bound (2.36), when the observable $f$ is a statistical estimator of an unknown deterministic parameter in a family of probability measures. It is also known, [12], that such bounds are sharp in the sense that for specific estimators (observables), e.g., the Maximum Likelihood Estimator, the bound (2.36) becomes an equality.

In Section 3.3 we obtain a new Cramer-Rao type inequality for time series based on our UQ Information bounds.

**3. Path-space UQ information inequalities and sensitivity bounds.** In this section we develop new uncertainty quantification information inequalities and related sensitivity bounds for stochastic processes and their path-dependent observables. The approach developed in the previous section is applicable to obtaining similar bounds for functionals of Markov processes, when combined with path-space Information Theory tools such as the Relative Entropy Rate (RER) and the path Fisher Information Matrix (pFIM). These concepts which are discussed next were introduced as UQ and sensitivity analysis tools for stochastic processes in [23, 24, 11].
3.1. Information theory metrics in path space. We consider stochastic processes which are Markov and take values in Polish space $X$, although a much more general set up is also possible, see for instance [17]. For simplicity in the presentation, we further restrict our discussion to discrete-time Markov processes $\{X_t\}_{t \in \mathbb{N}}$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ with the transition kernel $p(x, dy)$ and with the initial measure $\mu(dx)$, and the Markov process $\{Y_t\}_{t \in \mathbb{N}_0}$ with the transition kernel $q(x, dy)$ and with the stationary measure $\nu(dx)$. For the time interval $0, 1, ..., T$, we denote by $P_{[0,T]}$, $Q_{[0,T]}$ the respective probability measures on path space. Similar notation and constructions for all concepts introduced here will also be used when $t \in [0, \infty)$, we refer to the Appendix A, as well as to [17, 23].

We will assume conditions under which the path-space relative entropy

$$
\mathcal{R} (Q_{[0,T]} \| P_{[0,T]})
$$

is finite for all $T < \infty$. For stationary Markov processes, the relative entropy scales linearly in $T$ as $T \to \infty$, [17]. Thus it is natural to define the concept of the Relative Entropy Rate,

**Definition 3.1.** Let $P_{[0,T]}$ and $Q_{[0,T]}$ be path-measures corresponding to Markov processes $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$. We define the relative entropy rate by

$$
\mathcal{H}(Q \| P) = \lim_{T \to \infty} \frac{1}{T} \mathcal{R} (Q_{[0,T]} \| P_{[0,T]}) ,
$$

when the limit exists.

Although RER is a quantity between path distributions, we drop the dependence of time interval in the notation of the RER because RER is a time-independent quantity. Moreover, the relative entropy rate can often be expressed explicitly, which we demonstrate via examples in Appendix A. For instance, in the case of discrete-time Markov Chains we have

$$
\mathcal{H}(Q \| P) = \int_X \nu(dx) \int_X q(x, dy) \log \frac{dq(x, \cdot)}{dp(x, \cdot)}(y) = \int_X \mathcal{R} (q(x, \cdot) \| p(x, \cdot)) \nu(dx) .
$$

The significance of the definition of RER is elucidated by the following property of the relative entropy of two path-measures for stationary processes. We state it for simplicity in the case of discrete-time Markov Chains. For the proof we refer to Appendix A. The proof was first given by Shannon in [25] and since then has been extended in various directions for Markov and semi-Markov processes, [17].

**Lemma 3.2.** Let $\{X_t\}_{t \in \mathbb{N}_0}$, $\{Y_t\}_{t \in \mathbb{N}_0}$ be two stationary Markov chains with the path-measures $P_{[0,T]}$ and $Q_{[0,T]}$ and stationary measures $\mu$ and $\nu$. Then for any $T \in \mathbb{N}_0$

$$
\mathcal{R} (Q_{[0,T]} \| P_{[0,T]}) = T \mathcal{H}(Q \| P) + \mathcal{R} (\nu \| \mu) ,
$$

and the relative entropy rate $\mathcal{H}(Q \| P)$ is independent of $T$ and given by (3.3).

As in Section 2.5, we will consider the sensitivity analysis problem, but this time in the context of stationary dynamics. This amounts to an asymptotic expansion of the relative entropy, and eventually the RER, in terms of a parameter perturbation. First we consider the path-space probability measure $P^\theta_{[0,T]}$ where $\theta \in \mathbb{R}^k$ is a vector of the model parameters. We consider a perturbation $v \in \mathbb{R}^k$ in the parameter vector $\theta$ and the resulting path-space probability measure $P^\theta+v_{[0,T]}$. We start out with a regularity condition on the dependence of the probabilities on the parameter $\theta$; this condition is the weakest possible, but is simple to state.

**Condition 3.1.** There is a fixed reference probability measure $R \in \mathcal{P}(X)$ such that $P^\theta (x, dy) \ll R(dy)$ for all $x \in X$ and $\theta \in \mathbb{R}^k$. Let $p^\theta (x, y) = \frac{dP^\theta (x, \cdot)}{dR (\cdot)}(y)$. Then we assume $(x, y, \theta) \to p^\theta (x, y)$ is continuous and for each fixed $x, y$ that $\theta \to p^\theta (x, y)$ is $C^3$. We assume the same regularity for the stationary measures $\mu^\theta$, i.e., $\mu^\theta \ll R$ and $\theta \to \frac{d\mu^\theta}{dR} (x)$ is $C^4$ for each fixed $x$. Where needed, we also assume the existence of suitable dominating functions for various functions of $p^\theta$.

Note that under this assumption, the stationary distribution $\mu^\theta$ is absolutely continuous with respect to $R$. It also holds that $P^\theta_{[0,T]}$ is absolutely continuous with respect to the product measure on $X^T$ with
Concentrating on the stationary regime, the path-space relative entropy scales linearly with time as Lemma 3.2 and using the variational representation (2.7) of the cumulant-generating function, we obtain for any observable $\mathcal{F}$ where $\mu$ is a measurable functional $\mathcal{F}$ with $\mu$ stationary measure $\mu$ for any bounded observable function $f$. The right hand side of the equation scales most linearly as $T \to \infty$ and its correct rescaling for large times is given by

$$\frac{1}{T} \tilde{\Lambda}_{[0,T]}(c) = \sup_{Q_{[0,T]} \in \mathcal{P}_{[0,T]}} \left\{ -\mathcal{R}(Q_{[0,T]} || P_{[0,T]}) + c \mathbb{E}_{Q_{[0,T]}}[\mathcal{F}] - \mathbb{E}_{P_{[0,T]}}[\mathcal{F}] \right\}.$$ 

Concentrating on the stationary regime, the path-space relative entropy scales linearly with time as Lemma 3.2 shows. Moreover, if we consider observables for which $\mathbb{E}_{P_{[0,T]}}[\mathcal{F}]$ and $\mathbb{E}_{Q_{[0,T]}}[\mathcal{F}]$, are uniformly bounded for all $T$, then the second term in the supremum in (3.7) scales also linearly with time, therefore, the right hand side of the equation scales most linearly as $T \to \infty$ and its correct rescaling for large times is given by

$$\frac{1}{T} \tilde{\Lambda}_{[0,T]}(c) = \sup_{Q_{[0,T]} \in \mathcal{P}_{[0,T]}} \left\{ -\mathcal{R}(Q_{[0,T]} || P_{[0,T]}) + c \mathbb{E}_{Q_{[0,T]}}[\mathcal{F}] - \mathbb{E}_{P_{[0,T]}}[\mathcal{F}] \right\}.$$ 

One class of such observables which have a finite expectation as $T \to \infty$ is the case where $\mathcal{F}$ is bounded by a constant. Another type of observables of this category which is of great interest in stochastic computing is ergodic averages:

$$\mathcal{F}(X) = \frac{1}{T} \int_0^T f(X_s) ds,$$ 

for any bounded observable function $f$. Under proper ergodic assumptions one has

$$\lim_{T \to \infty} \mathcal{F}(X) = \mathbb{E}_{\mu}[f] = \int f d\mu.$$ 

Next, we provide a result on path space which is similar to (2.11), first obtained by in [6, 16] for measures $P$ and $Q$. We use the notation and the goal-oriented divergence formulation in Theorem 2.9. We show that
for (suitable) path-space observables, the analogue of relative entropy in (2.11) is now the concept of Relative Entropy Rate (3.1).

**Theorem 3.3.** Assume the conditions of Lemma 3.2 and that the time-averaged cumulant generating function

$$
\frac{1}{T} \tilde{\Lambda}_{P, T} f(c) \equiv \frac{1}{T} \log \mathbb{E}_{P_{0, T}}[e^{c T f - \mathbb{E}_{P_{0, T}}[f]})]
$$

exists in a neighborhood of the origin. Using the notation of Lemma 3.2 and the representation of the (path) relative entropy rate (3.1), we have

$$
\Xi_+(Q_{0, T} \mid P_{0, T}; F) \equiv \inf_{c > 0} \left\{ \frac{1}{T} \tilde{\Lambda}_{P, T} f(c) + \frac{1}{c} \left( \mathcal{H}(Q \mid P) + \frac{1}{T} \mathcal{R}(\nu \mid \mu) \right) \right\}, \tag{3.11}
$$

$$
\Xi_-(Q_{0, T} \mid P_{0, T}; F) \equiv \sup_{c > 0} \left\{ -\frac{1}{T} \tilde{\Lambda}_{P, T} f(-c) - \frac{1}{c} \left( \mathcal{H}(Q \mid P) + \frac{1}{T} \mathcal{R}(\nu \mid \mu) \right) \right\}. \tag{3.12}
$$

Then we have the bounds

$$
\Xi_-(Q_{0, T} \mid P_{0, T}; F) \leq \mathbb{E}_{Q_{0, T}}[F] - \mathbb{E}_{P_{0, T}}[F] \leq \Xi_+(Q_{0, T} \mid P_{0, T}; F). \tag{3.13}
$$

Furthermore, based on the divergence properties and analytic formula in Theorem 2.4 and Theorem 2.7 respectively for the divergences, we have

$$
\Xi_\pm(Q_{0, T} \mid P_{0, T}; F) = \tilde{\Lambda}_r f(c) = F(c) := -\tilde{\Lambda}_{P, T} f(c) + c \tilde{\Lambda}_{P, T} f(c)
$$

is a strictly increasing function on $(0, c)$, where $c = \sup \{ c : \tilde{\Lambda}_{P, T} f(c) < \infty \}$.

**Proof.** The proof follows immediately from Theorem 2.4 and Theorem 2.7 as well as the bounds (2.11) and the relative entropy rate representation of the relative entropy in Lemma 3.2, e.g. (3.1).

The bounds in Theorem 3.3 become (asymptotically) more explicit in the case where the relative entropy rate $\mathcal{H}(Q \mid P)$ is small, that is by expanding $\Xi_-(Q_{0, T} \mid P_{0, T}; F)$ and $\Xi_+(Q_{0, T} \mid P_{0, T}; F)$ in $\mathcal{H}(Q \mid P)$ in (3.13) in $\mathcal{H}(Q \mid P)$. Furthermore, the relative entropy rate $\mathcal{H}(Q \mid P)$ can be explicitly calculated in several examples discussed earlier in Section 3 and in Appendix A. More specifically we have the following asymptotics:

**Lemma 3.4.** Assume that the cumulant generating function $\Lambda_{P_{0, T}, T} f(c)$ exists in a neighborhood of the origin. Assume also that

$$
\frac{1}{T} \mathcal{R}(Q_{0, T} \mid P_{0, T}) = \rho^2
$$

for two path probability measures $P_{0, T}, Q_{0, T}$. [Note that by (3.10), this is essentially an assumption on the relative entropy rate $\mathcal{H}(Q \mid P)$.] Then, there exists a function $c_T(\rho)$ which is the unique minimizer (resp. maximizer) of (3.11) (resp. (3.12)). Furthermore, there is $\rho_0 > 0$ such that $c_T(\rho)$ is $C^\infty$ in $(0, \rho_0)$ and admits the perturbation expansion

$$
c_T(\rho) = c_{T, 1}^* \rho + O(\rho^2), \quad \text{where} \quad c_{T, 1}^* = \sqrt{\frac{2}{T \text{Var}_{P_{0, T}}[T F]}}.
$$

(3.15)

For observables of the form $F(X) = \frac{1}{T} \sum_{k=0}^{T-1} f(X_k)$, it holds that

$$
\tau_T(f) := \frac{1}{T} \text{Var}_{P_{0, T}}[T F] = \text{Var}_\mu[f] + 2 \sum_{k=1}^{T} \left( 1 - \frac{|k|}{T} \right) A_f(k).
$$

(3.16)
Where \( A_f(t) := E_{P_{0,T}}[(f(X_0) - E_{\mu}[f(X_0)])(f(X_t) - E_{\mu}[f(X_0)])] \) is the stationary covariance function.

Proof. The proof of all save \((3.16)\) follows the same steps as the proof of Lemma 2.11. The proof of \((3.16)\) is carried out in Lemma 3.8 below, see \((3.20)\).

Substituting the expansion in \( \rho \) for the optimal value \((3.15)\) into the expansion of \( \Xi_{\nu}(Q_{[0,T]} || P_{[0,T]}; F) \) we obtain asymptotics of the upper and lower bounds for the weak error in \( \rho^2 = \frac{1}{T} R(Q_{[0,T]} || P_{[0,T]}) \).

Theorem 3.5 (Linearization). Under the assumptions of Theorem 3.3, we have:

\[
|E_{Q_{[0,T]}[F]} - E_{P_{[0,T]}[F]}| \leq \sqrt{\frac{1}{T} \text{Var}_{P_{0,T}}[TF]} \sqrt{2(H(Q || P) + \frac{1}{T} R(\nu || \mu) + O\left(\frac{1}{T} R(Q_{[0,T]} || P_{[0,T]})\right)),
\]

where we recall \((3.5)\)

\[
\frac{1}{T} R\left(Q_{[0,T]} || P_{[0,T]}\right) = H(Q || P) + \frac{1}{T} R(\nu || \mu).
\]

As in the static case, the term \( O\left(\frac{1}{T} R(Q_{[0,T]} || P_{[0,T]})\right) \) in \((3.17)\) can be further quantified using the asymptotic expansions of \( c^*(\rho) \) and \((3.11)\) in \( \rho \).

3.3 Path-space perturbation analysis and sensitivity bounds. As in Section 2.5, we will consider the sensitivity analysis problem, but this time in the context of stationary dynamics. First we consider the path-space probability measure \( P^\theta_{[0,T]} \) where \( \theta \in \mathbb{R}^k \) is a vector model parameters. We consider a perturbation \( v \in \mathbb{R}^k \) in the parameter vector \( \theta \) and the resulting path-space probability measure \( P_{[0,T]}^{\theta+v} \) and focus first on the discrete time model. The continuous time calculations are carried out in a similar manner, and we refer to Appendix A for the related formulas.

Lemma 3.6. Assume the conditions of Lemma \((3.6)\) and Condition \((3.7)\). Assume that the time-averaged cumulant generating function \( \frac{1}{T} \tilde{\Lambda}^{\nu}_{P^\theta_{[0,T]}}, T_F(c) \equiv \frac{1}{T} \log E_{P^\theta_{[0,T]}}[e^{cT(F - E_{P^\theta_{[0,T]}}[F])}] \) exists in a neighborhood of the origin. Then for \( \nu \in \mathbb{R}^k \) and \( \epsilon \) in a neighborhood of the origin there exists a function \( c^*_F(\epsilon) \) which is the unique solution of both optimization problems \((3.11)\) and \((3.12)\), where \( P_{[0,T]} = P^\theta_{[0,T]} \) and \( Q_{[0,T]} = P^{\theta+\epsilon}_{[0,T]} \).

This function admits the perturbation expansion \( c^*_F(\epsilon) = c^*_{F,1}\epsilon + O(\epsilon^2) \), where

\[
c^*_{F,1} = \sqrt{\frac{E_{P^\theta_{[0,T]}}[TF]}{\text{Var}_{P^\theta_{[0,T]}}[TF]}}.
\]

Proof. The proof follows along the same steps as the proof of Lemma \((3.16)\) using also Lemma \((3.6)\) and \((3.6)\).

Then the next theorem readily follows from \((3.17)\).

Theorem 3.7 (Finite/Long-time regimes). Assume the conditions of Lemma \((3.6)\) For any \( \nu \in \mathbb{R}^k \) and \( \epsilon \neq 0 \),

\[
\frac{1}{|\epsilon|} |E_{P^\theta_{[0,T]}+[\epsilon]}[F] - E_{P^\theta_{[0,T]}}[F]| \leq \sqrt{\frac{1}{T} \text{Var}_{P^\theta_{[0,T]}}[TF]} \sqrt{v^T(I_H(P^\theta) + \frac{1}{T} I(\mu^\theta))v + O(\epsilon)},
\]

and

\[
|S_{F,\nu}(P^\theta_{[0,T]})| \leq \sqrt{\frac{1}{T} \text{Var}_{P^\theta_{[0,T]}}[TF]} \sqrt{v^T(I_H(P^\theta) + \frac{1}{T} I(\mu^\theta))v},
\]

where the sensitivity index \( S_{F,\nu}(P^\theta_{[0,T]}) \) is defined in \((2.31)\).

Remark 3.1. In the steady states regime and for time-averaged observables such as \((3.9)\), it holds that

\[
S_{F,\nu}(P^\theta_{[0,T]}) = S_{F,\nu}(\mu^\theta).
\]
where $\mu^\theta$ is the stationary distribution. The implications of this equivalence are further discussed in Section 3.5 and demonstrated in Section 4.

**Remark 3.2.** We note that the bound (3.19) relates gradient-type sensitivity analysis methods such as likelihood-ratio method [9] that develop efficient estimators for the sensitivity indices [2,3,7], with information-theory based methods, showing that the latter provide a sensitivity bound on (2.31). Similarly the bound (3.19) relates sensitivity methods relying on finite-differencing [1] with information-theory sensitivity analysis methods, [23]. We refer to the inequalities (3.19) and (3.29) as Sensitivity Bounds. These bounds can be computed efficiently and can provide fast screening of insensitive observables, as well as parameters or directions in the parameter space. We refer to [3] for more details, implementations and examples on this perspective.

Next we focus on the infinite-time asymptotic regime and the related Sensitivity Bounds. Taking the limit $T \to \infty$ we obtain bounds for ergodic-type observables.

**Lemma 3.8.** Under the assumptions of Lemma 3.6 and for observables of the form $f X \theta [0,T]$, the following conclusions hold. The limit $c^*_1 = \lim_{T \to \infty} c^*_{T,1}$ exists, and

$$
C^*_1 = \frac{\sum_{i,j} T H(P^\theta)^{i,j} v_i v_j}{\tau(f)},
$$

where $\tau(f)$ is the Integrated Autocorrelation Function (IAT), [12], defined as

$$
\tau(f) := \lim_{T \to \infty} \tau_T(f) = \text{Var}_{\mu^\theta}[f] + 2 \sum_{k=1}^{\infty} A_f(k),
$$

and

$$
A_f(k) := \mathbb{E} P^\theta [0,T] [(f(X_0) - \mathbb{E}_{\mu^\theta}[f(X_0)])(f(X_k) - \mathbb{E}_{\mu^\theta}[f(X_0)])]
$$

is the stationary covariance function of the process $X$.

**Proof.** Using that under $P^\theta [0,T]$ each $X_i$ is distributed according to $\mu^\theta$, a direct computation of the time-averaged variance gives

$$
\frac{1}{T} \text{Var}_{P^\theta [0,T]} (T F_T(X)) = \frac{1}{T} \mathbb{E} P^\theta [0,T] \left[ \left( \sum_{i=0}^{T-1} f(X_i) - \mathbb{E}_{P^\theta [0,T]} \left( \sum_{i=0}^{T-1} f(X_i) \right) \right)^2 \right]
$$

$$
= \frac{1}{T} \sum_{i=0}^{T-1} \sum_{j=0}^{T-1} \mathbb{E} P^\theta [0,T] \left[ \left( f(X_i) - \mathbb{E}_{P^\theta [0,T]} f(X_i) \right) \left( f(X_j) - \mathbb{E}_{P^\theta [0,T]} f(X_j) \right) \right]
$$

$$
= \frac{1}{T} \sum_{i=0}^{T-1} \sum_{j=0}^{T-1} \text{Cov}_f(i,j)
$$

where $\text{Cov}_f(i,j)$ is the covariance between $f(X_i)$ and $f(X_j)$. Owing to stationarity, we have that $\text{Cov}_f(i,j) = \mathbb{E} P^\theta [0,T] \left[ \left( f(X_i) - \mathbb{E}_{\mu^\theta}[f(X_0)] \right) \left( f(X_j) - \mathbb{E}_{\mu^\theta}[f(X_0)] \right) \right] = \text{Cov}_f(i-j,0) = A_f(i-j)$. Therefore,

$$
\frac{1}{T} \text{Var}_{P^\theta [0,T]} (T F_T(X)) = \frac{1}{T} \sum_{k=-T}^{T-1} (T - |k|) A_f(k,0) = \sum_{k=-\infty}^{\infty} \left( 1 - \frac{|k|}{T} \right) A_f(k)
$$

Sending $T \to \infty$, we obtain that

$$
\lim_{T \to \infty} \frac{1}{T} \text{Var}_{P^\theta [0,T]} (T F_T(X)) = \sum_{k=-\infty}^{\infty} A_f(k) = \tau(f)
$$
between $f$ of the stationary measure corresponding to the relative entropy. Finally, the static sensitivity bound (2.36) is the FIM of the transition probability function while in (2.36) arises the classical FIM of the stationary measure corresponding to the relative entropy. Furthermore, the second term in (3.29) is the pFIM corresponding to the relative time-autocorrelation sum of the observable. Therefore, the time-correlation crucially affects the behavior of the observable. Indeed, upon rearranging, this bound is precisely the sensitivity bound (2.36), where the expected value of the observable $f$ is the biased estimator of an unknown deterministic parameter in a family of probability measures. Specifically, assume a biased estimator $\hat{\theta}$ for the sensitivity indices defined by (2.31).

Upon rearranging, this bound is precisely the sensitivity bound (2.36), where the expected value of the observable $f$ is the biased estimator of an unknown deterministic parameter in a family of probability measures. Specifically, assume a biased estimator $\hat{\theta} = f(X)$ of the parameter $\theta$ with bias function $\psi(\theta)$, i.e., $\mathbb{E}_{f}(\hat{\theta}) = \psi(\theta)$. Then the Cramer-Rao bound for a scalar parameter $\theta$ states that

$$\text{Var}_{f}(\hat{\theta}) \geq \frac{[\psi'(\theta)]^2}{\mathcal{I}(\theta)}.$$

(3.30)

Upon rearranging, this bound is precisely the sensitivity bound (2.36), where the expected value of the observable $f$ is the biased estimator of an unknown deterministic parameter in a family of probability measures. Furthermore, it is also known, [12, 13], that such bounds are sharp in the sense that for specific estimators (observables), e.g. the Maximum Likelihood Estimator, the bound (3.30), (2.36) becomes an equality.

In the same sense, we can obtain a new Cramer-Rao type inequality for time series stationary statistics based on our UQ Information bounds in path-space. Indeed, path-space observables such as $\mathcal{F}_{T}(X) = \frac{1}{T} \sum_{i} f(X_{i})$ play the role of the statistical estimator for $\theta$, (i.e., $\hat{\theta} = \mathcal{F}_{T}(X)$) and the sensitivity bound (3.29) constitutes a Cramer-Rao lower bound for the IAT (3.24),

$$\tau_{f}(\theta) \geq \frac{[\psi'(\theta)]^2}{\mathcal{I}(\theta)}.$$

(3.31)

where $\psi(\theta) = \mathbb{E}_{\theta} [\mathcal{F}_{T}]$ is the bias of the estimator. Therefore, for dependent samples created for instance by Monte Carlo Markov Chain methods [12], the lower bound (3.31) can be utilized. Finally, we remark that estimators with dependent samples have generally larger variance than estimators using independent samples, however, for the same amount of computational time, larger number of dependent samples than independent samples are drawn. Hence it is not clear which estimator has better performance in terms of reduced variance for a given computational cost. In this direction, the Cramer-Rao bound (3.31) may be very useful.

### 3.4. Cramer-Rao inequalities for time-series

The sensitivity bounds (3.20) and (3.29) can be considered as extension of the Cramer-Rao inequality for the time-series of Markov processes. Indeed, we recall that for a parametric family of probability measures $P^{\theta}$, where for simplicity in the presentation we assume that $\theta$ is scalar, the Cramer-Rao inequality provides a lower bound for the variance of any unbounded statistical estimator. Specifically, assume a biased estimator $\hat{\theta} = f(X)$ of the parameter $\theta$ with bias function $\psi(\theta)$, i.e., $\mathbb{E}_{f}(\hat{\theta}) = \psi(\theta)$. Then the Cramer-Rao bound for a scalar parameter $\theta$ states that

$$\text{Var}_{f}(\hat{\theta}) \geq \frac{[\psi'(\theta)]^2}{\mathcal{I}(\theta)}.$$

(3.30)

Upon rearranging, this bound is precisely the sensitivity bound (2.36), where the expected value of the observable $f$ is the biased estimator of an unknown deterministic parameter in a family of probability measures. Furthermore, it is also known, [12, 13], that such bounds are sharp in the sense that for specific estimators (observables), e.g. the Maximum Likelihood Estimator, the bound (3.30), (2.36) becomes an equality.

In the same sense, we can obtain a new Cramer-Rao type inequality for time series stationary statistics based on our UQ Information bounds in path-space. Indeed, path-space observables such as $\mathcal{F}_{T}(X) = \frac{1}{T} \sum_{i} f(X_{i})$ play the role of the statistical estimator for $\theta$, (i.e., $\hat{\theta} = \mathcal{F}_{T}(X)$) and the sensitivity bound (3.29) constitutes a Cramer-Rao lower bound for the IAT (3.24),

$$\tau_{f}(\theta) \geq \frac{[\psi'(\theta)]^2}{\mathcal{I}(\theta)}.$$

(3.31)

where $\psi(\theta) = \mathbb{E}_{\theta} [\mathcal{F}_{T}]$ is the bias of the estimator. Therefore, for dependent samples created for instance by Monte Carlo Markov Chain methods [12], the lower bound (3.31) can be utilized. Finally, we remark that estimators with dependent samples have generally larger variance than estimators using independent samples, however, for the same amount of computational time, larger number of dependent samples than independent samples are drawn. Hence it is not clear which estimator has better performance in terms of reduced variance for a given computational cost. In this direction, the Cramer-Rao bound (3.31) may be very useful.

### 3.5. Discussion

Even though the form of the sensitivity bounds in this section are similar to (2.36) there are some substantial differences. For example, the first term of the path-space bound (3.29) is the time-autocorrelation sum of the observable. Therefore, the time-correlation crucially affects the behavior of the sensitivity bound. Furthermore, the second term in (3.29) is the pFIM corresponding to the relative entropy rate (i.e., the FIM of the transition probability function) while in (2.36) arises the classical FIM of the stationary measure corresponding to the relative entropy. Finally, the static sensitivity bound (2.36)
can be obtained from (3.29) if we draw independent samples from the stationary measure. Next, we explore some of these issues in different directions.

**Remark 3.4.** When we want to study the sensitivity of ergodic averages such as (3.22), we can either use the path space estimate in (3.29)

$$|S_{f,v}(\mu^\theta)| \leq \sqrt{\tau(f)} \sqrt{\sum_{ij} I_H(P^\theta)_{ij} v_i v_j}, \quad (3.32)$$

or alternatively the equilibrium bound (2.36), namely,

$$|S_{f,v}(\mu^\theta)| \leq \sqrt{\text{Var}_{\mu^\theta}[f]} \sqrt{\sum_{ij} I(\mu^\theta)_{ij} v_i v_j}. \quad (3.33)$$

On one hand, (3.33) involves the FIM of the equilibrium measures $\mu^\theta$, which we do not typically have available in most non-equilibrium systems such as biochemical networks, reaction-diffusion mechanisms or driven systems. However the pathwise estimate (3.32) can in principle always be computed since it involves only the local dynamics $p^\theta$ in the path Fisher Information Matrix (3.5). Finally, for more general observables $\mathcal{F}$ than (3.9) or (3.22), the path-wise sensitivity bounds (3.32) are the only ones available.

**Remark 3.5.** Given the results in this section, as well as the computational feasibility of RER and path FIM demonstrated in [23, 11], it is reasonable to investigate the class of functionals of the path $\{X_t\}$ that can be controlled by these quantities. Based on (3.17), it readily follows that we need to consider path-space observables $\mathcal{F}$ such that for some constant $C$,

$$\frac{1}{T} \text{Var}_{P^{\theta}[0,T]}(\mathcal{F}) = T \text{Var}_{P^{\theta}[0,T]}(\mathcal{F}) \leq C < \infty, \quad \text{uniformly in } T.$$  

More general conditions of similar nature can be also derived based on the finiteness of the upper and bounds in (3.19), as $T \to \infty$.

**Remark 3.6.** Let $\hat{f}(l)$ be the discrete-time Fourier transform of $\{f(X_k)\}_{k=0}^\infty$, then the Fourier transform of the stationary autocorrelation function, $A_f(k)$, is the well-known power spectral density [4] which equals to $|\hat{f}(l)|^2$. Thus, IAT equals $\tau(f) = |\hat{f}(0)|^2$ and the path-space sensitivity bound (3.32) can be rewritten as

$$|S_{f,v}(\mu^\theta)| \leq |\hat{f}(0)| \sqrt{I_H(P^\theta)_{ij} v_i v_j} \quad (3.34)$$

4. Demonstration Examples. This section demonstrates the application of the derived bounds for several stochastic models. The sensitivity bound derived in Section 2.1 are utilized in the first two examples where the sharpness of the bound is discussed. In the third and forth examples, both stationary and path space bounds are derived and compared for various observable functions. In these examples, the stationary bounds are slightly sharper than the bounds that utilize the pathwise FIM, however, stationary bounds are rarely explicitly available. Indeed, the birth/death process presented in Section 4.3 is a special case of a single-species biochemical reaction network with explicit stationary distribution, however, for reaction networks with more species stationary distribution is hardly known. Similarly, the stationary distribution in Section 4.4 where a stochastic differential equation example is considered is not generally known when the drift term is not of conservative type (i.e., the gradient of an appropriate function). For such stochastic models, the only possibly available option for a tractable sensitivity bound is the pathwise bound.

4.1. Exponential family of distributions. A probability density function belongs to the exponential family if it admits the following canonical decomposition [21]

$$P^\theta(x) = \exp\{t(x)^T \theta - F(\theta) + k(x)\}$$

where $t(x) = [t_1(x), ..., t_K(x)]^T$ is the sufficient statistics vector, $\theta \in \mathbb{R}^K$ is the parameter vector, $F(\cdot)$ is the log-normalizer (log-partition function or free energy in statistical physics) and $k(x)$ is the carrier measure.
Statistics $t(x)$ was named sufficient after Sir R. Fisher in 1922 because it contains all the information needed for the estimation of the parameters. Considering the sufficient statistics as observables, the corresponding sensitivity indices can be analytically calculated as

$$S_{tk,p^o}(\theta_l) = \left[ \frac{\partial^2}{\partial\theta_k \partial\theta_l} \mathbb{E}_{p^o}[t_k(x)] \right] = \left[ \frac{\partial^2}{\partial\theta_k \partial\theta_l} F(\theta) \right], \quad k, l = 1, ..., K$$

The covariance matrix of the sufficient statistics vector equals the Hessian of the log-normalizer, $F$, (i.e., $\text{Cov}_{p^o}(t(x)) = \nabla^2 F(\theta)$) while the relative entropy of $P^\theta$ w.r.t. $P^{\theta+\epsilon}$ can be written as the Bregman divergence of the log-normalizer on swapped natural parameters [21] given by

$$\mathcal{R}(P^\theta || P^{\theta+\epsilon}) = F(\theta + \epsilon) - F(\theta) - \epsilon^T \nabla F(\theta).$$

A straightforward Taylor series expansion of $F$ in $\epsilon$ implies that the Fisher information matrix, $\mathcal{I}(P^\theta)$, defined in (2.30) equals the Hessian of the log-normalizer, too.

Therefore, for sufficient statistics of the exponential family distribution, Theorem 2.14 states that

$$\left| S_{tk,p^o}(\theta_l) \right| \leq \sqrt{\text{Var}_{p^o}(t_k) I(P^\theta)_{l,l}} \leq \sqrt{\left[ \frac{\partial^2}{\partial\theta_k^2} F(\theta) \right] \left[ \frac{\partial^2}{\partial\theta_l^2} F(\theta) \right]}.$$

Notice that the inequality becomes an equality when $k = l$. From a parameter estimation perspective, the equality of the bound of the $k$-th sufficient statistic with respect to the $k$-th parameter is equivalent to the fact that $t_k(x)$ is an efficient estimator of $\theta_k$, $k = 1, ..., K$. In other words, the Cramer-Rao bound is attained [15, Thm 5.12]. Finally, another bound for the sensitivity indices can be obtained from the properties of the sufficient statistics vector as well as the initial data $u$.

4.2. Uncertainty quantification example. Let’s consider the differential equation

$$\dot{u} = - Xu, \quad u(0) = u_0$$

where $X$ is a Gaussian random variable with mean $\mu$ and variance $\sigma^2$. This stochastic model has been previously utilized for the assessment of uncertainty quantification bounds in [16]. The stochastic solution of the equation is

$$u(t) = u_0 e^{-Xt}$$

whose distribution law is log-normal with parameters $\log(u_0) - \mu t$ and $(\sigma t)^2$. The probability density function is given at time instant $t$ by

$$P^\theta(u) = \frac{1}{u\sigma t \sqrt{2\pi}} \exp \left\{ - \frac{(\log(u) - \log(u_0) + \mu t)^2}{2(\sigma t)^2} \right\}$$

where $\theta = [\mu, \sigma]^T$ while the dependence of the probability function on time $t$ as well as the initial data $u_0$ is hidden for the sake of notational simplicity. The goal is to compute the observable that quantifies the probability of $u(t)$ being larger that a determined value, $\bar{u}$, at time instant $t$. This is a failure probability and can be written as an ensemble average,

$$P_f = \mathbb{E}_{p^\theta} \left[ \chi_{\{u > \bar{u}\}} \right],$$

where the observable function is the characteristic function (i.e., $f(u) = \chi_{\{u > \bar{u}\}}(u)$). Notice that even though log-normal distribution belongs to the exponential family, we are neither interested in the natural parameters nor in the sufficient statistics. Therefore, the general setting of the previous subsection does not apply.
Nevertheless, detailed calculation can be performed for this example. Indeed, the computation of the sensitivity index for $\mu$ is given by

$$S_{\chi(\{u > \bar{u}\})}(\mu) = -\mathbb{E}_{P^\theta}[\chi(\{u > \bar{u}\})(u) \frac{\log(u) - \log(u_0) + \mu t}{\sigma^2 t}],$$

while the sensitivity index for the standard deviation $\sigma$ is

$$S_{\chi(\{u > \bar{u}\}),P^\theta}(\sigma) = \mathbb{E}_{P^\theta}\left[\chi(\{u > \bar{u}\})(u) \frac{(\log(u) - \log(u_0) + \mu t)^2 - \sigma^2 t^2}{\sigma^3 t^2}\right].$$

The variance of the observable is

$$\text{Var}_{P^\theta}(\chi(\{u > \bar{u}\})) = \mathbb{E}_{P^\theta}[\chi(\{u > \bar{u}\})^2] - (\mathbb{E}_{P^\theta}[\chi(\{u > \bar{u}\})])^2 = \frac{1}{4} - \frac{1}{4} \text{erf}^2 \left(\frac{\log(\bar{u}) - \log(u_0) + \mu t}{\sqrt{2} \sigma t}\right)$$

where erf is the error function while the diagonal elements of the Fisher information matrix for the log-normal distribution, $P^\theta$, are given by

$$I(P^\theta)_{1,1} = \mathbb{E}_{P^\theta}\left[\frac{(\log(u) - \log(u_0) + \mu t)^2}{(\sigma^2 t)^2}\right]$$

and

$$I(P^\theta)_{2,2} = \mathbb{E}_{P^\theta}\left[\frac{(\log(u) - \log(u_0) + \mu t)^2 - \sigma^2 t^2}{(\sigma^3 t^2)^2}\right].$$

Figures 4.1 and 4.2 show the absolute value of the sensitivity indices and the corresponding sensitivity bounds as a function of time for $\sigma = 1$ and $\sigma = 2$, respectively. The remaining parameters were set to $u_0 = 1$, $\bar{u} = 10$ and $\mu = 1$ while the computations of the expectations were performed numerically, whenever necessary. In Figure 4.1 the sensitivity bound of Theorem 2.14 follows closely the sensitivity index at the course of time. The sensitivity bound in Figure 4.2 performs accurately for the sensitivity index of the mean (upper panel), however, it is less sharp around time $t = 5$ for the standard deviation (lower panel) due to the existence of a zero transition of the sensitivity index at that particular instant. Interestingly enough, the lower panel of Figure 4.2 reveals that both upper and lower bounds for small time and larger times, respectively, provide information about the corresponding sensitivity index. Overall, taking into account the complexity of the chosen observable which can be a risk-sensitive (i.e., rare event) observable the performance of the sensitivity bound is considered satisfactory.

### 4.3. Birth/death process

We consider a well-mixed reaction network which consists of one species and two reactions given by

$$\emptyset \xrightarrow{k_1} X \xleftarrow{k_2} \emptyset.$$

The corresponding propensity functions for the current state $x = x$ are

$$a_1(x) = k_1 \quad \text{and} \quad a_2(x) = k_2 x.$$

Mathematically, this stochastic system is modelled as a continuous-time Markov chain (CTMC) and due to its simplicity there exist analytic representations of the steady state (equilibrium) distribution, moments and autocorrelation function. [8, Sec. 7.1]. The steady state distribution, $\mu^\theta$, of the reaction network is Poisson with parameter $\frac{k_1}{k_2}$. Hence, the steady state moments as well as the FIM for the parameter vector $\theta = [k_1, k_2]^T$ are known. The elements of the stationary FIM (eq. (2.30)) are shown in Table 4.1. In the same Table, the elements of the pathwise FIM are shown [24, pp. 10]. Notice that the stationary FIM is singular while the pathwise FIM is full rank implying that when the complete time-series is provided then both parameters can be inferred. If samples were i.i.d. drawn from the steady state distribution, then only the ratio is inferable.
Proceeding, let us consider two observables; the mean, $f_1(x) = x$, and, the variance, $f_2(x) = (x - \frac{k_1}{k_2})^2$. Since $E_{\mu^e}[f_1] = E_{\mu^s}[f_2] = \frac{k_1}{k_2}$, the sensitivity indices are $S_{f_1,\mu^e}(k_1) = S_{f_2,\mu^e}(k_1) = \frac{1}{k_2}$ and $S_{f_1,\mu^e}(k_2) = S_{f_2,\mu^e}(k_2) = -\frac{k_1}{k_2^2}$. Moreover, in order to compute the IAT for $f_1$ and $f_2$, the computation of the autocorrelation and the autocorrelation of the variance are necessary. Due to the linear nature of this example\[8\], explicit formulae exist and they are reported in Table 4.2. The corresponding IATs are also shown in Table 4.2 for both observable functions.

In Table 4.3 both stationary and pathwise sensitivity bounds are compared with the actual sensitivity indices. The fact that Poisson distribution belongs to the exponential family results in sharp bound for the
mean value and the stationary case while the bound for the pathwise case is worse by a $\sqrt{2}$ factor. When the variance is considered as an observable, the stationary bound is also slightly tighter than the pathwise bound. In the extreme cases for the variance observable, the pathwise bound becomes equivalent to the stationary bound when $k_2 \ll k_1$ while both bounds become sharper when $k_1 \ll k_2$. Finally, note that even though the comparable performance of both stationary and pathwise bounds, there is a crucial advantage of the pathwise analysis which is its computational tractability. Indeed, in complex reaction networks, the steady state distribution is rarely known, hence, the stationary FIM cannot be derived. On the other hand, explicit formula for the pathwise FIM exists \cite{24} and the corresponding sensitivity bound is computable.

**Table 4.2**

Variances, autocorrelation function and IAT for the observables $f_1(x)$ and $f_2(x)$ of the birth/death process.

| Observable | Variance | ACF | IAT |
|------------|----------|-----|-----|
| $f_1(x) = x$ | $\frac{k_1}{k_2}$ | $\frac{k_1}{k_2} e^{-k_2|t|}$ | $\frac{2k_1}{k_2}$ |
| $f_2(x) = (x - \frac{k_1}{k_2})^2$ | $\frac{k_1}{k_2} + 2\frac{k_1^2}{k_2^2}$ | $\frac{k_1}{k_2} e^{-k_2|t|} + 2\frac{k_1^2}{k_2^2} e^{-2k_2|t|}$ | $2\left(\frac{k_1}{k_2} + \frac{k_1^2}{k_2^2}\right)$ |

4.4. Ornstein-Uhlenbeck process. Consider a one-dimensional Ornstein-Uhlenbeck (OU) process defined by the following stochastic differential equation

$$dX_t = -\alpha (X_t - \beta) dt + \gamma dB_t$$

where $X_t$ is the process, $\theta = [\alpha, \beta, \gamma]^T$ are the system’s parameters while $B_t$ is a one-dimensional Brownian motion. The drift of the OU process is a linear function of the state and the noise is additive since the diffusion term does not depend on the state. The stationary distribution of the OU process, $\mu^o$, is Gaussian with mean $\beta$ and variance $\frac{\gamma^2}{2\alpha}$. The diagonal elements of the stationary FIM are thus calculated and they are presented in Table 4.3 (2nd column). Taking $f(x) = x$ as an observable, Table 4.4 reports the variance with respect to the stationary measure, the autocorrelation function as well as the IAT for the continuous-time process.

There are two basic approaches for the computation of the pathwise FIM. The first is to compute RER directly from the Girsanov formula and then FIM is obtained from a linearization procedure. The formula for RER is given in (A.9), thus, it is straightforward to calculate the pathwise FIM whose diagonal elements are shown in Table 4.5. Notice that if the diffusion parameter, $\gamma$, is perturbed by a small amount then the RER is infinite. Indeed, Girsanov theorem stated that the path-space measures of two SDE processes

| SI | Value | SB (Thm 4.4) | SB (Thm 4.5) |
|----|-------|-------------|-------------|
| $S_{x,\mu^o}(k_1)$ | $\frac{1}{k_2}$ | $\frac{1}{k_2}$ | $\frac{\sqrt{2}}{k_2}$ |
| $S_{x,\mu^o}(k_2)$ | $- \frac{k_1}{k_2}$ | $\frac{1}{k_2}$ | $\frac{\sqrt{2}}{k_2}$ |
| $S_{(x - \frac{k_1}{k_2})^2,\mu^o}(k_1)$ | $\frac{1}{k_2}$ | $\frac{1}{k_2} \sqrt{1 + 2\frac{k_1}{k_2}}$ | $\frac{\sqrt{2}}{k_2} \sqrt{1 + 2\frac{k_1}{k_2}}$ |
| $S_{(x - \frac{k_1}{k_2})^2,\mu^o}(k_2)$ | $- \frac{k_1}{k_2}$ | $\frac{k_1}{k_2} \sqrt{1+2\frac{k_1}{k_2}}$ | $\frac{\sqrt{2}}{k_2} \sqrt{1 + 2\frac{k_1}{k_2}}$ |
be not absolute continuous w.r.t. each other when the diffusion terms are different. Therefore, the pathwise sensitivity bound in continuous-time is applicable only for the parameters of the drift. The second approach is to discretize the stochastic process, defining a new discrete-time Markov chain and then compute the pathwise FIM from the FIM of the DTMC renormalized with the time-step. Even though the second approach is an approximation, it admits an important advantage which is the capability of providing a sensitivity bound even when the diffusion parameters are considered. The trick of time-discretization resulted in regularization of the new path-space measures, hence, a finite RER is obtained even if the parameters of the diffusion part are perturbed.

### Table 4.4

| Observable | Variance | ACF (cont. time) | IAT (cont. time) | ACF (Euler) | IAT (Euler) |
|------------|----------|-----------------|-----------------|-------------|-------------|
| $f(x) = x$ | $\frac{\gamma}{2}$ | $\frac{\gamma}{2} e^{-\alpha t}$ | $\frac{\gamma}{2}$ | $\frac{\gamma}{2} (1 - \alpha \Delta t)^2$ | $\frac{\gamma}{2}$ |

Proceeding with the second approach and for the OU process, we consider the Euler scheme which is a first-order weak error integrator [14] given at the $n$-th step by

$$X_{n+1} = X_n + \alpha(X_n - \beta) \Delta t + \gamma \sqrt{\Delta t} \Delta W_n$$

where $\Delta t$ is the discretization step while $\Delta W_n$ are i.i.d. zero-mean Gaussians with unit variance. Hence, the transition probability, $p^0(x, y)$, is Gaussian with mean $x + \alpha(x - \beta) \Delta t$ and variance $\gamma^2 \Delta t$. The last two columns of Table 4.4 shows the autocorrelation function as well as the IAT for the discrete-time process obtained after discretization using Euler scheme while the last column of Table 4.3 shows the diagonal elements of the pathwise FIM again for the same discrete-time process. In order to compute these quantities, averaging with respect to the (unknown) stationary distribution of the Euler scheme, $\bar{\mu}$, which is an approximation of the stationary distribution of the continuous-time process, $\mu^0$, is required. However, we averaged with respect to $\mu^0$ instead of $\bar{\mu}$ exploiting the fact that the produced weak error is of order $O(\Delta t)$, [20]. Another remark on the pathwise FIM is that when the limit $\Delta t \to 0$ is taken and the diffusion parameter, $\gamma$, is perturbed then the corresponding FIM value is infinite which is in accordance with the Girsanov theorem.

### Table 4.5

| Order | Stationary | Pathwise (cont. time) | Pathwise (Euler) |
|-------|------------|-----------------------|-----------------|
| $I(1, 1)$ | $\frac{1}{\sqrt{2\pi}}$ | $\frac{1}{\sqrt{2\pi}}$ | $\frac{1}{\sqrt{2\pi}}$ |
| $I(2, 2)$ | $\frac{\alpha}{\sqrt{\pi}}$ | $\frac{\alpha}{\sqrt{\pi}}$ | $\frac{\alpha}{\sqrt{\pi}}$ |
| $I(3, 3)$ | $\frac{\gamma}{\sqrt{\pi}}$ | $\infty$ | $\frac{\gamma}{\sqrt{\pi}\Delta t}$ |

Table 4.6 presents the sensitivity indices and the various sensitivity bounds for the mean value as an observable. Stationary bound for $\beta$ is sharp as expected due to the fact that Gaussian belongs to the exponential family and mean value is a sufficient statistic. Continuous-time pathwise bound as well as discrete-time pathwise bound (up to order $O(\Delta t)$) for $\beta$ are sharp. For $\alpha$, the stationary bound is smaller by a factor of $\sqrt{2}$ while for $\gamma$ the factor $\frac{\sqrt{2}}{\Delta t}$ of the discrete-time pathwise bound make the stationary bound better. Finally, notice that as in the birth/death process the stationary bounds are slightly tighter. However, for general SDEs where the drift term is not necessarily of conservative type, the stationary distribution is rarely known hence the computation of stationary FIM and consequently the stationary bounds are intractable. For instance, a large class of stochastic processes where the stationary distribution is hardly known is the non-equilibrium systems where the drift consists of non-conservative forces. Therefore, the respective stationary sensitivity bound is intractable for this important category of stochastic processes.

**Appendix A. Relative entropy rate and path Fisher information matrix: Examples.** The relative entropy rate (RER) and the path Fisher Information Matrix (pFIM) can often be expressed explicitly...
Furthermore, the relative entropy rate is expressed as the relative entropy
\[ \mu \] where processes with values in the Polish space \( X \) discrete and continuous time Markov Chains and Stochastic Differential Equations.

A.1. Discrete time Markov chains. RER always has an explicit expression for discrete time processes with values in the Polish space \( X \). We first state a version of the chain rule. For a proof see [7 Theorem C.3.1].

**Lemma A.1.** Let \( \alpha \) and \( \beta \) be probability measures on \( X \times Y \), where \( X \) and \( Y \) are Polish spaces. Let \( \alpha_1 \) and \( \beta_1 \) denote their first marginals, and denote by \( \alpha(dy|x) \) and \( \beta(dy|x) \) the conditional distribution on the second variable given the first. Then the mapping \( x \rightarrow \mathcal{R}(\alpha(\cdot|x) || \beta(\cdot|x)) \) is measurable, and

\[
\mathcal{R}(\alpha \| \beta) = \mathcal{R}(\alpha(\cdot|x) || \beta(\cdot|x)) + \int_X \mathcal{R}(\alpha(\cdot|x) || \beta(\cdot|x)) \alpha_1(dx).
\]

**Lemma A.2.** Let \( \{X_t\}_{t \geq 0}, \{Y_t\}_{t \geq 0} \) where \( t \in \mathbb{N}_0 \), be Markov processes on the state space \( X \) with transition kernels \( p(x, dx') \) and \( q(x, dx') \), and initial measures \( \mu(dx) \) and \( \nu(dx) \), respectively. Assume that \( \mu \) is stationary for \( p(x, dx') \). Then the relative entropy rate \( \mathcal{H}(P || Q) \) defined in (3.1) is given by

\[
\mathcal{H}(P || Q) = \int \mu(dx) \int p(x, dy) \log \frac{dp(x, \cdot)}{dq(x, \cdot)}(y). \tag{A.1}
\]

Furthermore, the relative entropy rate is expressed as the relative entropy

\[
\mathcal{H}(P || Q) = \mathcal{R}(\mu \otimes p || \mu \otimes q), \tag{A.2}
\]

where \( \mu \otimes p \) is the probability measure on \( X^2 \) given by \( [\mu \otimes p](A \times B) = \int_A p(x, B) \mu(dx). \)

**Proof.** Both statements follow directly from the chain rule. Since \( \mu \) is stationary for \( p(x, dx') \), we can apply the chain rule from time \( t = T - 1 \) back to \( t = 0 \), and by using Markov property obtain (A.2), with \( \mathcal{H}(P || Q) \) equal to

\[
\int_X \mathcal{R}(p(x, \cdot) || q(x, \cdot)) \mu(dx),
\]

which is precisely (A.1). (A.2) also follows directly from the chain rule and the fact that \( \mathcal{R}(\mu || \mu) = 0 \). Finally, notice that even though a quantity between path distributions, we drop the dependence of time interval in the notation of the relative entropy rate because relative entropy rate is a time-independent quantity. \( \square \)

**Lemma A.3.** Assume Condition (3.1). Then, the path FIM defined in (3.4) is given by

\[
\mathcal{I}_H(P^\theta) = \mathbb{E}_{\mu^\theta} \left[ \int_E p^\theta(x, y) \nabla_\theta \log p^\theta(x, y) \nabla_\theta \log p^\theta(x, y)^T R(dy) \right]. \tag{A.3}
\]

**Proof.** Define the function \( G(\theta) = G(\theta; x, y) = \log p^\theta(x, y) \) for all \( x, y \in X \). Then, from Condition 3.1 \( G(\theta) \) as a function of \( \theta \) is \( C^3 \) and for an arbitrary \( \epsilon \in \mathbb{R}^k \)

\[
G(\theta + \epsilon) = G(\theta) + \epsilon^T \nabla_\theta G(\theta) + \frac{1}{2} \epsilon^T \nabla_\theta^2 G(\theta) \epsilon + R_2(\theta)
\]

\[
= G(\theta) + \epsilon^T \nabla_\theta p^\theta \frac{\nabla_\theta p^\theta}{p^\theta} + \frac{1}{2} \epsilon^T \left( \frac{\nabla_\theta p^\theta}{p^\theta} \right)^2 \epsilon + R_2(\theta),
\]
where \( \nabla \) and \( \nabla^2 \) denotes the gradient and the Hessian of a function while \( R_2(\theta) \) is the remainder term as given by Taylor’s Theorem.

Then, the relative entropy rate of the path distribution \( P_{[0,T]}^\theta \) with respect to the perturbed path distribution \( P_{[0,T]}^{\theta+\epsilon} \) becomes

\[
\mathcal{H}(P^\theta \| P^{\theta+\epsilon}) = \int \mu^\theta(dx) \int p^\theta(x,y) \log \frac{p^\theta(x,y)}{p^{\theta+\epsilon}(x,y)} R(dy)
\]

\[
= - \int \mu^\theta(dx) \int p^\theta(x,y)(G(\theta + \epsilon; x, y) - G(\theta; x, y)) R(dy)
\]

\[
= - \int \mu^\theta(dx) \int p^\theta(x,y) \left( \epsilon \nabla p^\theta(x,y) \mu^\theta(dy) + \frac{1}{2} \epsilon^T \left( \frac{\nabla^2 p^\theta(x,y)}{p^\theta(x,y)} \right) \epsilon + R_2(\theta; x, y) \right) R(dy)
\]

\[
= \frac{1}{2} \epsilon^T \int \mu^\theta(dx) \int p^\theta(x,y) \left( \frac{\nabla^2 p^\theta(x,y)}{p^\theta(x,y)} \right)^2 R(dy) \epsilon + \int \mu^\theta(dx) \int p^\theta(x,y) R_2(\theta; x, y) R(dy)
\]

since for any \( i = 1, 2, \ldots \) it holds that

\[
\int p^\theta(x,y) \frac{\nabla_i p^\theta(x,y)}{p^\theta(x,y)} R(dy) = \int \nabla_i p^\theta(x,y) R(dy) = \nabla_i \int p^\theta(x,y) R(dy) = \nabla_i 1 = 0
\]

Thus, the path FIM is given by

\[
\mathcal{I}_\mathcal{H}(P^\theta) = \mathbb{E}_{\mu^\theta} \left[ \int p^\theta(x,y) \nabla_\theta \log p^\theta(x,y) \nabla_\theta \log p^\theta(x,y)^T R(dy) \right]
\]

**Remark A.1.** Performing similar Taylor series expansion, it can be obtained that the relative entropy rate of \( P^{\theta+\epsilon} \) w.r.t. \( P^\theta \) admits the same Hessian. Indeed, it is expanded as

\[
\mathcal{H}(P^{\theta+\epsilon} \| P^\theta) = \frac{1}{2} \epsilon^T \mathcal{I}_\mathcal{H}(P^\theta) \epsilon + O(|\epsilon|^3).
\]

Notice also that this result is valid not only for discrete-time Markov chains but it is quite general.

**A.2. Continuous-time Markov chains.** The next natural step is to compute the relative entropy rate for continuous-time Markov chains. We consider such chains on a countable state space \( \mathcal{X} \) and let quantities such as \( P_{[0,T]}^\theta \) denote the measure on \( D([0,T] : \mathcal{X}) \) induced by the process, where \( D([0,T] : \mathcal{X}) \) consists of all \( \xi : [0,T] \rightarrow \mathcal{X} \) that are continuous from the right and with limits from the left, with the usual Skorohod topology.

**Lemma A.4.** Let \( \{X_t\}_{t \geq 0} \) and \( \{Y_t\}_{t \geq 0} \) be stationary continuous time Markov chains with the countable state space \( \mathcal{X} \) and jump rates \( \lambda(x) \) and \( \tilde{\lambda}(x) \) and transition probabilities \( p(x,x') \) and \( \tilde{p}(x,x') \). Assume that \( \lambda \) and \( \tilde{\lambda} \) are positive and uniformly bounded above. Assume also that \( p(x,x) = \tilde{p}(x,x) = 0 \) for all \( x \in \mathcal{X} \), and for \( x \neq y \) that \( p(x,y) > 0 \) if any only if \( \tilde{p}(x,y) > 0 \). Let \( \mu \) be a stationary probability distribution for \( \{X_t\}_{t \geq 0} \), and let \( \tilde{\mu} \) be any initial distribution for \( \{Y_t\}_{t \geq 0} \). Let \( P_{[0,T]} \) and \( Q_{[0,T]} \) be the measures induced by \( \{X_t\}_{t \geq 0} \) and \( \{Y_t\}_{t \geq 0} \). Then the relative entropy rate \( \mathcal{H}(P \| Q) \) associated with \( \mathcal{R}(P_{[0,T]} \| Q_{[0,T]}) \) is given by

\[
\mathcal{H}(P \| Q) = \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \mu(x)\lambda(x)p(x,x') \log \frac{\lambda(x)p(x,x')}{\tilde{\lambda}(x)\tilde{p}(x,x')} - \sum_{x \in \mathcal{X}} \mu(x)(\lambda(x) - \tilde{\lambda}(x)). \tag{A.4}
\]

**Proof.** According to [13 Prop. 2.6, App. 1] and [18 Sec. 19] the Radon Nikodym derivative pf the path measure \( P_{[0,T]} \) with respect to the path measure \( Q_{[0,T]} \) is given by

\[
\frac{dP_{[0,T]}(\xi)}{dQ_{[0,T]}(\xi)} = \frac{\mu(\xi_0)}{\tilde{\mu}(\xi_0)} \exp \left( \int_0^T \log \frac{\lambda(x)p(x,x')}{\tilde{\lambda}(x)\tilde{p}(x,x')} dN_t(x) - \int_0^T \frac{\lambda(x) - \tilde{\lambda}(x)}{\tilde{\mu}(\xi_t) - \tilde{\lambda}(x)} dt \right),
\]
where \( N_t(\xi) \) is the number of jumps on the path \( \xi \) up to time \( s \). The relative entropy up to time \( T \) is defined by

\[
R(P_{[0,T]} \mid Q_{[0,T]}) \equiv E_{P_{[0,T]}} \left[ \log \frac{dP_{[0,T]}}{dQ_{[0,T]}} \right].
\]

Since \( \lambda \) is bounded \( M_T \equiv N_T - \int_0^T \lambda(\xi_t) \, dt \) is a mean zero martingale, then for any (non-negative and measurable) function \( f \) on \( \mathcal{X} \)

\[
E_{P_{[0,T]}} \left[ \int_0^T f(\xi_t) \, dN_t \right] = E_{P_{[0,T]}} \left[ \int_0^T f(\xi_t) \lambda(\xi_t) \, dt \right].
\]

Furthermore, from the stationarity of the process, we have

\[
E \left[ \sum_{x \in \mathcal{X}} \mu(x) \log \frac{\lambda(x)p(x,x')}{\lambda(x')p(x,x')} - \sum_{x \in \mathcal{X}} \mu(x)(\lambda(x) - \tilde{\lambda}(x)) \right] + R(\mu \mid \tilde{\mu}).
\]

**Remark A.2.** One can rearrange the expression for the RER to obtain

\[
\mathcal{H}(P \mid Q) = \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} \mu(x) \tilde{\lambda}(x)p(x,x') \ell \left( \frac{\lambda(x)p(x,x')}{\lambda(x')p(x,x')} \right),
\]

where

\[
\ell(z) = z \log z - z + 1 \quad \text{for } z \geq 0.
\]

This exhibits the RER as a form of relative entropy. The function \( \ell(z) \), which appears in rate functions for the large deviation theory of jump Markov processes \cite{touchette2009}, is non-negative and vanishes only at \( z = 1 \). Thus the RER is nonnegative, and equals zero if and only if the two chains are the same.

**Lemma A.5.** Let the transition rate defined for all \( x, x' \in \mathcal{X} \) by \( e^0(x,x') \equiv \lambda^0(x) p^0(x,x') \) be parametrized by \( \theta \in \mathbb{R} \) and assume that the mapping \( \theta \to e^0(\cdot, \cdot) \) is \( C^2 \). Let \( P_{[0,T]}^\theta \) (resp. \( \mu^\theta \)) be the path (resp. stationary) measure of the associated process. Then, the path FIM is

\[
\mathcal{I}_{\mathcal{H}}(P^\theta) = E_{\mu^\theta} \left[ \sum_{x \in \mathcal{X}} e^\theta(x, x') \nabla_\theta \log e^\theta(x, x') \nabla_\theta \log e^\theta(x, x')^T \right].
\]

**Proof.** The proof is similar to the DTMC case using now two auxiliary functions defined by \( G_1(\theta) = G_1(\theta; x, x') = \log e^\theta(x, x') \) and \( G_2(\theta) = G_2(\theta; x, x') = e^\theta(x, x') \) for all \( x, x' \in \mathcal{X} \). For completeness, we present the basic steps of the relative entropy expansion. The relative entropy rate of the path measure \( P_{[0,T]}^\theta \) with respect to the perturbed path measure \( P_{[0,T]}^{\theta + \epsilon} \) can be written as

\[
\mathcal{H}(P^\theta \mid P^{\theta + \epsilon}) = \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) e^\theta(x, x') \log \frac{e^\theta(x, x')}{e^{\theta + \epsilon}(x, x')} - \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) (e^\theta(x, x') - e^{\theta + \epsilon}(x, x'))
\]

\[
= - \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) e^\theta(x, x')(G_1(\theta + \epsilon) - G_1(\theta)) + \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) (G_2(\theta + \epsilon) - G_2(\theta))
\]

\[
= - \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) e^\theta(x, x') \left( e^T \nabla \sigma e^\theta(x, x') + \frac{1}{2} e^T \nabla_{\sigma}^2 e^\theta(x, x') - \left( \nabla \sigma e^\theta(x, x') \right)^2 \right) + R_2(\theta; x, x')
\]

\[
+ \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) \left( e^T \nabla \sigma e^\theta(x, x') + \frac{1}{2} e^T \nabla_{\sigma}^2 e^\theta(x, x') \right) + \tilde{R}_2(\theta; x, x')
\]

\[
= \frac{1}{2} \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) e^\theta(x, x') \left( \nabla \sigma e^\theta(x, x') \right)^2 - \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) (e^\theta(x, x') R_2(\theta; x, x') - \tilde{R}_2(\theta; x, x'))
\]

\[
= \frac{1}{2} \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) e^\theta(x, x') \left( \nabla \sigma e^\theta(x, x') \right)^2 - \sum_{x, x' \in \mathcal{X}} \mu^\theta(x) (e^\theta(x, x') R_2(\theta; x, x') - \tilde{R}_2(\theta; x, x'))
\]
where $R_{2}(\theta)$ and $\hat{R}_{2}(\theta)$ are the remainder terms of $G_{1}$ and $G_{2}$, respectively.

**A.3. Stochastic differential equations.** We also compute the relative entropy rate for Ito diffusion processes. To avoid technical difficulties we impose following assumptions: we assume that the vector fields $a(x), b(x) \in \mathbb{R}^{d}$, $x \in \mathbb{R}^{d}$ and the non-singular $\sigma(x) \in \mathbb{R}^{d \times d}$ are such that the Ito’s stochastic differential equations

$$
dX_t = a(X_t)dt + \sigma(X_t)dW_t, \quad (A.6)$$
$$
dY_t = b(Y_t)dt + \sigma(Y_t)dW_t, \quad (A.7)$$

have a unique weak solution for initial conditions $X_0 \sim \mu_0(dx)$ and $Y_0 \sim \nu_0(dx)$. We also assume that the solutions $\{X_t\}_{t \geq 0}$ and $\{Y_t\}_{t \geq 0}$ are ergodic and we denote $\mu(dx)$ and $\nu(dx)$ the respective invariant measures. Furthermore, we assume that the function

$$u(x) = \sigma^{-1}(x)(b(x) - a(x))$$

is such that Novikov’s condition $\mathbb{E} \left[ e^{\frac{1}{2} \int_0^T |u(X_t)|^2 dt} \right] < \infty$ is satisfied. Under these assumptions we obtain explicit formula for the relative entropy rate of the stationary process $\{X_t\}_{t \geq 0}$ that is the solution of (A.6) with the initial condition $X_0 \sim \mu(dx)$, where $\mu(dx)$ is the invariant distribution.

**Lemma A.6.** Let $\{X_t\}_{t \geq 0}$ and $\{Y_t\}_{t \geq 0}$ be the unique solutions of (A.6)-(A.7) with the initial conditions $X_0 \sim \mu_0(dx)$ and $Y_0 \sim \nu_0(dx)$, where $\mu_0(dx) = \mu(dx)$ is the invariant distribution for the process $\{X_t\}_{t \geq 0}$. We define $u(x) = \sigma^{-1}(x)(a(x) - b(x))$. Denoting $P_{[0,T]}$ and $Q_{[0,T]}$ the corresponding path probability measures, the relative entropy is

$$\mathcal{R}(P_{[0,T]} || Q_{[0,T]}) = \mathbb{E}_{P_{[0,T]}} \left[ \frac{1}{2} \int_0^T |u(X_t)|^2 dt \right] + \mathcal{R}(\mu_0 || \nu_0), \quad (A.8)$$

and the relative entropy rate $\mathcal{H}(P || Q) \equiv \lim_{T \to \infty} \frac{1}{T} \mathcal{R}(P_{[0,T]} || Q_{[0,T]})$ is

$$\mathcal{H}(P || Q) = \mathbb{E}_{\mu} \left[ \frac{1}{2} \|a - b\|_{\Sigma^{-1}}^2 \right], \quad (A.9)$$

where $\|b\|_{\Sigma^{-1}} \equiv \sum_{i,j=1}^{d} \Sigma_{ij}^{-1}(x)b_i(x)b_j(x)$ is the norm on $\mathbb{R}^d$ defined by the diffusion matrix $\Sigma = \sigma(x)\sigma^T(x)$.

**Proof.** Under the assumptions on the stochastic differential equations it follows from Girsanov’s Theorem, [22], that $Q_{[0,T]} \ll P_{[0,T]}$ and the Radon-Nikodym derivative is

$$
\frac{dP_{[0,T]}(X_t)}{dQ_{[0,T]}(X_t)} = \frac{d\mu_0}{d\nu_0}(X_0) e^{-\int_0^T u(X_s) dW_s - \frac{1}{2} \int_0^T |u(X_s)|^2 ds}.
$$

Furthermore, $B_t = \int_0^t u(X_s) ds + W_t$ is Brownian motion under $P_{[0,T]}$. Thus we have

$$\mathcal{R}(P_{[0,T]} || Q_{[0,T]}) = \mathbb{E}_{P_{[0,T]}} \left[ \log \frac{dP_{[0,T]}(X_t)}{dQ_{[0,T]}(X_t)} \right]$$
$$= \mathcal{R}(\mu_0 || \nu_0) + \mathbb{E}_{P_{[0,T]}} \left[ -\int_0^T u(X_s) dW_s - \frac{1}{2} \int_0^T |u(X_s)|^2 ds \right]$$
$$= \mathcal{R}(\mu_0 || \nu_0) + \mathbb{E}_{P_{[0,T]}} \left[ -\int_0^T u(X_s) dB_s - u(X_s) ds - \frac{1}{2} \int_0^T |u(X_s)|^2 ds \right]$$
$$= \mathcal{R}(\mu_0 || \nu_0) + \mathbb{E}_{P_{[0,T]}} \left[ \frac{1}{2} \int_0^T |u(X_s)|^2 ds \right],$$

where in the last identity we use $\mathbb{E}_{P_{[0,T]}} \left[ -\int_0^T u(X_s) dB_s \right] = 0$ as $B_t$ is Brownian motion under $P_{[0,T]}$. If $X_0 \sim \mu$ and thus the process $\{X_t\}_{t \geq 0}$ is stationary we have

$$\mathbb{E}_{P_{[0,T]}} \left[ \frac{1}{2} \int_0^T |u(X_s)|^2 ds \right] = T \mathbb{E}_{\mu} \left[ \frac{1}{2} |u(x)|^2 \right],$$
Lemma A.7. Let the drift term $a^\theta(x)$ be parametrized by $\theta \in \mathbb{R}$ and assume that the mapping $\theta \to a^\theta(\cdot)$ is $C^2$. Let $P^\theta_{[0,T]}$ (resp. $\mu^\theta$) be the path (resp. stationary) measure of the associated process. Then, the path Fisher Information Matrix (FIM) is

$$\mathcal{I}_\mathcal{H}(P^\theta) = \mathbb{E}_{\mu^\theta} \left[ \nabla_{\theta a^\theta}(x) (\sigma \sigma^T)^{-1}(x) \nabla_{\theta a^\theta}(x) \right].$$

(A.10)

Proof. Taylor’s theorem for the drift term $a^\theta(\cdot)$ around $\theta$ reads

$$a^{\theta + \epsilon}(x) = a^\theta(x) + \nabla_{\theta a^\theta}(x) \epsilon + R_1(\theta),$$

where $\nabla_{\theta a^\theta}(\cdot)$ is a $d \times k$ matrix containing all the first-order partial derivatives of the drift vector (i.e., the Jacobian matrix) while the vector $R_1(\theta)$ is the remainder term of the Taylor’s theorem. Then, the relative entropy rate of the path probability measure $P^\theta_{[0,T]}$ with respect to the perturbed path probability measure $P^{\theta + \epsilon}_{[0,T]}$ can be written as

$$\mathcal{H}(P^\theta \| P^{\theta + \epsilon}) = \frac{1}{2} \mathbb{E}_{\mu^\theta} \left[ \left| \sigma^{-1}(x) (a^{\theta + \epsilon}(x) - a^\theta(x)) \right|^2 \right]$$

$$= \frac{1}{2} \mathbb{E}_{\mu^\theta} \left[ (\nabla_{\theta a^\theta}(x) \epsilon + R_1(\theta; x))^T (\sigma \sigma^T)^{-1}(x) (\nabla_{\theta a^\theta}(x) \epsilon + R_1(\theta; x)) \right]$$

$$= \frac{1}{2} \mathbb{E}_{\mu^\theta} \left[ \nabla_{\theta a^\theta}(x) \epsilon (\sigma \sigma^T)^{-1}(x) \nabla_{\theta a^\theta}(x) \epsilon \right]$$

$$+ \epsilon^T \mathbb{E}_{\mu^\theta} \left[ \nabla_{\theta a^\theta}(x)^T (\sigma \sigma^T)^{-1}(x) R_1(\theta; x) \right] + \frac{1}{2} \mathbb{E}_{\mu^\theta} \left[ |\sigma^{-1}(x) R_1(\theta; x)|^2 \right]$$

from which (A.10) follows.
28] F. Nielsen and V. Garcia, Statistical exponential families: A digest with flash cards, arXiv.org:0911.4863, (2009).
22] B. Oksendal, Stochastic Differential Equations: An introduction with applications, Springer-Verlag, 2000.
23] Y. Pantazis and M. Katsoulakis, A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics, J. Chem. Phys., 138 (2013), p. 054115.
24] Y. Pantazis, M. Katsoulakis, and D. Vlachos, Parametric sensitivity analysis for biochemical reaction networks based on pathwise information theory, BMC Bioinformatics, 14:311 (2013).
25] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal, 27 (1948), pp. 379–423.
26] A. Tsybakov, Introduction to Nonparametric Estimation, Springer, 2008.