Uncertainty Quantification for Markov Processes via Variational Principles and Functional Inequalities

Jeremiah Birrell · Luc Rey-Bellet

Abstract Information-theory based variational principles have proven effective at providing scalable uncertainty quantification (i.e. robustness) bounds for quantities of interest in the presence of non-parametric model-form uncertainty. In this work, we combine such variational formulas with functional inequalities (Poincaré, log-Sobolev, Liapunov functions) to derive explicit uncertainty quantification bounds applicable to both discrete and continuous-time Markov processes. These bounds are well-behaved in the infinite-time limit and apply to steady-states.

Keywords uncertainty quantification · Markov process · relative entropy · Poincaré inequality · log-Sobolev inequality · Liapunov function · Bernstein inequality

Mathematics Subject Classification (2010) 47D07 · 39B72 · 60F10 · 60J25

1 Introduction

Information-theory based variational principles have proven effective at providing uncertainty quantification (i.e. robustness) bounds for quantities-of-interest in the presence of non-parametric model-form uncertainty [1,2,3,4,5,6,7,8,9,10]. In the present work, we combine these tools with functional

J. Birrell
Department of Mathematics and Statistics,
University of Massachusetts Amherst
Amherst, MA, 01003, USA
E-mail: birrell@math.umass.edu

Luc Rey-Bellet
Department of Mathematics and Statistics,
University of Massachusetts Amherst
Amherst, MA, 01003, USA
E-mail: luc@math.umass.edu

inequalities to obtain improved and explicit uncertainty quantification (UQ) bounds for both discrete and continuous-time Markov processes on general state spaces.

In our approach we are given a baseline model, described by a probability measure P; this is the model one has 'in hand' and that is amenable to analysis/simulation, but it may contain many sources of error and uncertainty. Perhaps it depends on parameters with uncertain values (obtained from experiment, Monte-Carlo simulation, variational inference, etc.) or is obtained via some approximation procedure (dimension reduction, neglecting memory terms, asymptotic approximation, etc.) In short, any quantity of interest computed from P has (potentially) significant uncertainty associated with it. Mathematically we chose to express this uncertainty by considering a (non-parametric) family, $\mathcal{U}(P)$, of alternative models that we postulate contains the inaccessible 'true' model.

Loosely stated, given some observable f, the uncertainty quantification goal considered here is

$$\textbf{Bound the bias } E_{\widetilde{P}}[f] - E_P[f] \text{ where } \widetilde{P} \in \mathcal{U}_r(P). \tag{1}$$

The subscript r indicates that the 'neighborhood' of alternative models, $\mathcal{U}_r(P)$, is often defined in terms of an error tolerance, $r > 0$. For our purposes, the appropriate notion of neighborhood will be expressed in terms of relative entropy, which can be interpreted as measuring the loss of information due to uncertainties. We do not discuss in full generality how to choose the tolerance level r but there are cases where one has enough information about the 'true' model to choose an appropriate tolerance; see Section 6.

Remark 1 Note that in Eq. (1), and the remainder of this paper, we consider the case where the quantity-of-interest is the expected value of some function, but extentions of these ideas to other quantities-of-interest are possible [6].

There are classical inequalities addressing Eq. (1) (ex: Csiszar-Kullback-Pinsker, Le Cam, Scheffé, etc.), but they exhibit poor scaling properties with problem dimension and/or in the infinite time limit. This problem is addressed by using tight information inequalities based on the Gibbs variational principle base and that are summarized in Section 2. See [5] for a detailed discussion of these issues.

The innovation of the present work is the use of functional inequalities in combination with the above mentioned variational approach to Eq. (1), thereby resulting in UQ bounds for Markov processes in the long-time regime. More specifically, given a continuous-time Markov process (X_t, P^μ) with initial distribution μ and stationary distribution μ^*, and an alternative (not-necessarily Markov) process $(X_t, \widetilde{P}^{\tilde{\mu}})$, we consider the problem of bounding the bias when the finite-time averages are computed by sampling from the alternative process:

$$\textbf{Bound} \qquad \widetilde{E}^{\tilde{\mu}}\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \int f d\mu^*. \tag{2}$$

Here, $\widetilde{E}^{\widetilde{\mu}}$ denotes the expectation with respect to $\widetilde{P}^{\widetilde{\mu}}$ and similarly for P^μ, E^μ. (Discrete time processes will also be considered in Section 5.)

Eq. (2) is a (much less studied) variant of the classical problem of the convergence of ergodic averages to the expectation in the stationary distribution:

$$E^\mu \left[\frac{1}{T} \int_0^T f(X_t)dt \right] \to \int f d\mu^*. \tag{3}$$

By combining information on the problems Eq. (2) and Eq. (3), one can also obtain bounds on the finite time sampling error:

$$\mathrm{err}_T = E^\mu \left[\frac{1}{T} \int_0^T f(X_t)dt \right] - \widetilde{E}^{\widetilde{\mu}} \left[\frac{1}{T} \int_0^T f(X_t)dt \right]. \tag{4}$$

Here we focus on the robustness problem, Eq. (2).

1.1 Summary of Results

The basis for all of our bounds is Theorem 3 in Section 2:

$$\pm \left(\widetilde{E}^{\widetilde{\mu}} \left[\frac{1}{T} \int_0^T f(X_t)dt \right] - \mu^*[f] \right) \le \inf_{c>0} \left\{ \frac{1}{cT} \Lambda^{\widehat{f}_T}_{P_T^{\mu^*}}(\pm c) + \frac{1}{cT} R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}) \right\}, \tag{5}$$

along with Corollary 2 in Section 3:

$$\frac{1}{T} \Lambda^{\widehat{f}_T}_{P_T^{\mu^*}}(\pm c) \le \kappa(V_{\pm c}), \tag{6}$$

$$\kappa(V) \equiv \sup \left\{ \langle A[g], g \rangle + \int V |g|^2 d\mu^* : g \in D(A, \mathbb{R}), \|g\|_{L^2(\mu^*)} = 1 \right\}, \tag{7}$$

$$V_{\pm c}(x) \equiv \pm c \left(f(x) - \mu^*[f] \right), \quad \mu^*[f] \equiv \int f d\mu^*. \tag{8}$$

In the above, $\Lambda^{\widehat{f}_T}_{P_T^{\mu^*}}(\pm c)$ is the cumulant generating function (see Eq. (35) for details), $R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*})$ is the relative entropy of the processes up to time T (see Eq. (15)), $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2(\mu^*)$, and $(A, D(A, \mathbb{R}))$ is the generator of the Markov semigroup for the process (X_t, P^μ) on $L^2(\mu^*)$.

Eq. (5) is derived by employing the Gibbs variational principle (hence the relation to relative entropy). Eq. (6), which is based on a theorem proven in [11], results from a connection between the cumulant generating function and the Feynman-Kac semigroup (hence the appearance of the generator, A). Also, note that the bound is expected to behave well in the limit $T \to \infty$, as $R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*})/T$ converges to the relative entropy rate of the processes, under suitable ergodicity assumptions.

Eq. (6) allows us employ our primary new tool for UQ, that being functional inequalities. By functional inequalities, we mean bounds on the generator, A, that will yield bounds on $\kappa(V_{\pm c})$; we will cover Poincaré, log-Sobolev, and F-Sobolev inequalities, as well as Liapunov functions. Our results rely heavily on the bounds obtained in [11,12,13,14,15] where concentration inequalities for ergodic averages where obtained.

The method outlined above will lead to explicit UQ bounds, expressed in terms of the observable, relative entropy, and the constants appearing in the functional inequalities (these constants are properties of the baseline process, P, only). The latter is potentially also a drawback, as computing explicit, tight constants for these functional inequalities is generally a very difficult problem. A second potential drawback of this approach is that most of these functional inequalities only involve the symmetric part of the generator (see Eq. (7)), hence the bounds are generally less than ideal for non-reversible systems.

For a simple example of the type of result obtained below, consider diffusion on $\mathbb{R}^n$ in a C^2 potential, V, i.e. the generator is $A = \Delta - \nabla V \cdot V$ and the invariant measure is $\mu^* = e^{-V(x)}dx$. Suppose the Hessian of V is bounded below:

$$D^2 V(x) \geq \alpha^{-1} I, \;\; \alpha > 0. \tag{9}$$

Our results give a Bernstein-type UQ bound for any bounded f:

$$\pm \left(\widetilde{E}^{\widetilde{\mu}} \left[\frac{1}{T} \int_0^T f(X_t)dt \right] - \mu^*[f] \right) \leq \sqrt{2\sigma^2 \eta} + M^{\pm}\eta, \tag{10}$$

$$M^{\pm} = \alpha \|(f - \mu^*[f])^{\pm}\|_{\infty}, \;\; \sigma^2 = 2\alpha \operatorname{Var}_{\mu^*}[f], \;\; \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu^*}).$$

See Sections 4.3.2 and 4.4.1 for further applications and references regarding diffusions.

The remainder of the paper is structured as follows. Necessary background on UQ for both general measures and processes will be given in Section 2, leading up to a connection with both the Feynman-Kac semigroup and relative entropy rate. Relevant properties of the Feynman-Kac semigroup are given in Section 3, culminating with the bound Eq. (6). The use of functional inequalities to obtain explicit bounds from Eq. (6) will be explored in Section 4. In Section 5 we show how these ideas can be adapted to discrete-time processes. Finally, the problem of bounding the relative entropy rate will be addressed in Section 6.

2 Uncertainty Quantification for Markov Processes

2.1 UQ via Variational Principles

Here we record important background information on the variational-principle approach to UQ.

Let P be a probability measure on a measurable space $(\Omega, \mathcal{F})$. We consider the class of random variables $f : \Omega \to \mathbb{R}$ with a well-defined and finite moment generating function:

$$\mathcal{E}(P) = \left\{ f : \Omega \to \mathbb{R} \; : \; E_P[e^{\pm c_0 f}] < \infty \text{ for some } c_0 > 0 \right\} . \qquad (11)$$

It is not difficult to prove (see e.g. [16]) that the cumulant generating function

$$\Lambda_P^f(c) = \log E_P[e^{cf}] \qquad (12)$$

is a convex function, finite and infinitely differentiable in some interval (c_-, c_+) with $-\infty \leq c_- < 0 < c_+ \leq \infty$ and equal to $+\infty$ outside of $[c_-, c_+]$. Moreover if $f \in \mathcal{E}(P)$ then f has moments of all orders and we write

$$\widehat{f} = f - E_P[f] \qquad (13)$$

for the centered observable of mean 0. We will often use the cumulant generating function for the centered observable $\widehat{f}$:

$$\Lambda_P^{\widehat{f}}(c) = \log E_P[e^{c(f - E_P[f])}] = \Lambda_P^f(c) - c E_P[f] . \qquad (14)$$

Recall also that the relative entropy (or Kullback-Leibler divergence) is defined by

$$R(\widetilde{P}||P) = \begin{cases} E_{\widetilde{P}}\left[\log \left(\frac{d\widetilde{P}}{dP} \right) \right] & \text{if } \widetilde{P} \ll P \\ +\infty & \text{otherwise} \end{cases} . \qquad (15)$$

It has the property of a divergence, that is $R(\widetilde{P}||P) \geq 0$ and $R(\widetilde{P}||P) = 0$ if and only if $\widetilde{P} = P$.

A key ingredient in our approach is the Gibbs Variational principle which relates the cumulant generating function and relative entropy; see [17].

Proposition 1 (Gibbs Variational Principle) *If $E_P[e^f] < \infty$ we have*

$$\log E_P[e^f] = \sup_{\widetilde{P} : R(\widetilde{P}||P) < \infty} \left\{ E_{\widetilde{P}}[f] - R(\widetilde{P}||P) \right\} , \qquad (16)$$

and the supremum is attained if and only if $\widetilde{P} = P^f$ is the tilted measure with $dP^f = e^{(f - \log E_P[e^f])} dP$.

As shown in [1,2], the Gibbs variational principle implies the following UQ bound for the expected values: (a similar inequality is used in the context of concentration inequalities see e.g. [18] and was also used independently in [8, 10]):

Theorem 1 (Gibbs information inequality) *If $R(\widetilde{P}||P) < \infty$ and $f \in \mathcal{E}(P)$ then $f \in L^1(\widetilde{P})$ and*

$$-\inf_{c>0} \left\{ \frac{\Lambda_P^{\widehat{f}}(-c) + R(\widetilde{P}||P)}{c} \right\} \leq E_{\widetilde{P}}[f] - E_P[f] \leq \inf_{c>0} \left\{ \frac{\Lambda_P^{\widehat{f}}(c) + R(\widetilde{P}||P)}{c} \right\} .$$
$$(17)$$

Theorem 1 is the basis for all further UQ bounds derived in this paper.

Remark 2 Note that even if $R(\widetilde{P}||P) = \infty$, the bound Eq. (17) trivially holds as long as $E_{\widetilde{P}}[f]$ is defined. To avoid clutter in the statement of our results, when $R(\widetilde{P}||P) = \infty$ we will consider the bound to be satisfied for any $f \in \mathcal{E}(P)$, even if $E_{\widetilde{P}}[f]$ is undefined.

Optimization problems of the form in Eq. (17) will appear frequently, hence we make the following definition:

Definition 1 Given any $\Lambda : \mathbb{R} \to [0, \infty]$ and $\eta > 0$, let

$$\Xi^{\pm}(\Lambda, \eta) \equiv \inf_{c>0} \left\{ \frac{\Lambda(\pm c) + \eta}{c} \right\}. \tag{18}$$

With this, we can rewrite the bound (17) as

$$-\Xi^{-}\left(\Lambda_P^{\widehat{f}}, R(\widetilde{P}||P)\right) \leq E_{\widetilde{P}}[f] - E_P[f] \leq \Xi^{+}\left(\Lambda_P^{\widehat{f}}, R(\widetilde{P}||P)\right). \tag{19}$$

2.2 Properties of $\Xi^{\pm}$

The objects

$$\Xi(\widetilde{P}||P; \pm f) \equiv \Xi^{\pm}\left(\Lambda_P^{\widehat{f}}, R(\widetilde{P}||P)\right) \tag{20}$$

appearing in the Gibbs information inequality, Eq. (19), have many remarkable properties, of which we list a few.

Theorem 2 *Assume $R(\widetilde{P}||P) < \infty$ and $f \in \mathcal{E}(P)$. We have:*

1. **(Divergence)** *$\Xi(\widetilde{P}||P; f)$ is a divergence, i.e. $\Xi(\widetilde{P}||P, f) \geq 0$ and $\Xi(\widetilde{P}||P; f) = 0$ if and only if either $P = \widetilde{P}$ or f is constant P a.s.*
2. **(Linearization)** *If $R(\widetilde{P}||P)$ is sufficiently small we have*

$$\Xi(\widetilde{P}||P, f) = \sqrt{2\mathrm{Var}_P[f]R(\widetilde{P}||P)} + O(R(\widetilde{P}||P)). \tag{21}$$

3. **(Tightness I)** *For $\eta > 0$ consider $\mathcal{U}_\eta = \{\widetilde{P}; R(\widetilde{P}||P) \leq \eta\}$. There exists η^* with $0 < \eta^* \leq \infty$ such that for any $\eta < \eta^*$ there exists a measure P^η with*

$$\sup_{\widetilde{P} \in \mathcal{U}_\eta} \left\{ E_{\widetilde{P}}[f] - E_P[f] \right\} = E_{P^\eta}[f] - E_P[f] = \Xi(P^\eta||P; f). \tag{22}$$

The measure P^η has the form

$$dP^\eta = e^{cf - \Lambda_P^f(c)}dP, \tag{23}$$

where $c = c(\eta)$ is the unique non-negative solution of $R(P^\eta||P) = \eta$.

Proof Items 1 and 2 are proved in [2]; see also [9] for item 2. Various version of the proof of item 3 can be found in in [1] or [2]. See Proposition 3 in [6] for a more detailed statement of the result; see also similar results in [8,19]. □

The tightness properties in Theorem 2 are very attractive and ultimately rely on the presence of the cumulant generating function $\Lambda_P^{\widehat{f}}(c)$, which encodes the entire law of f. However, this generally this makes the bound very difficult or impossible to compute explicitly; we will need to weaken Eq. (19) to obtain more usable bounds. Functional inequalities are one tool we will employ (see Section 4). Another ingredient, which we discuss next, will be explicit bounds on the optimization problem in the definition of $\Xi^{\pm}(\Lambda, \eta)$. Such an approach was put forward in [7] where various concentration inequalities such as Hoeffding, sub-Gaussian, and Bennett bounds are discussed. For this paper we will almost exclusively use the following Bernstein-type bound:

Lemma 1 *Suppose there exists $\sigma > 0$, $M^{\pm} \geq 0$ such that*

$$\Lambda(\pm c) \leq \frac{\sigma^2 c^2}{2(1 - cM^{\pm})} \tag{24}$$

for all $0 < c < 1/M^{\pm}$. Then for all $\eta \geq 0$ we have

$$\Xi^{\pm}(\Lambda, \eta) \leq \sqrt{2\sigma^2 \eta} + M^{\pm}\eta. \tag{25}$$

Note that $M^{\pm} = 0$ covers the case of a (one-sided) sub-Gaussian concentration bound.

Proof Bound Λ using Eq. (24) and solve the resulting optimization problem on $0 < c < 1/M^{\pm}$. □

From the point of view of concentration inequalities, the bound Eq. (24) is not very tight; indeed it holds for the cumulant generating function $\Lambda_P^{\widehat{f}}$ of any random variable $f \in \mathcal{E}(P)$, but explicit constants may be hard to come by. In the context of Markov process it has however been proved to be extremely useful, see [11,12,13,14] and in particular [15].

Second, we will need a linearization bound, generalizing Eq. (21):

Lemma 2 *Let $\Lambda : \mathbb{R} \to [0, \infty]$ be C^2 on a neighborhood of 0, $\Lambda(0) = \Lambda'(0) = 0$, and $\Lambda''(0) > 0$. Then*

$$\inf_{c>0} \left\{ \frac{\Lambda(\pm c) + \eta}{c} \right\} \leq \sqrt{2\Lambda''(0)\eta} + o(\sqrt{\eta}) \tag{26}$$

as $\rho \searrow 0$. If Λ'' is Lipschitz at 0 then the error bound improves of $O(\eta)$.

Proof The bound follows from Taylor expansion of $\Lambda(c)$; see the proof of Theorem 2.8 in [2]. □

2.3 UQ for Markov Processes

One of the main advantages of the Gibbs information inequality, Eq. (1), over classical information inequalities (such as Kullback-Leibler-Cziszàr inequality) is how it scales with time when applied to the distributions of processes on path space. See [5] for a detailed discussion of this issue. This strength will become apparent as we proceed.

The following assumption details the setting in which we will work for the remainder of this paper:

Assumption 1 *Let $\mathcal{X}$ be a Polish space and suppose we have a time homogeneous, $\mathcal{X}$-valued, càdlàg Markov family $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P^x)$, $x \in \mathcal{X}$, with transition probability kernel p_t. More specifically:*

1. *$(\Omega, \mathcal{F}, \mathcal{F}_t)$, $t \geq 0$ is a filtered probability space and X_t is a $\mathcal{X}$-valued, $\mathcal{F}_t$-adapted, càdlàg process.*
2. *$p_t(x, dy)$, $t \geq 0$, are time homogeneous transition probabilities on $\mathcal{X}$.*
3. *P^x, $x \in \mathcal{X}$ are probability measures with $(X_0)_* P^x = \delta_x$ for each $x \in \mathcal{X}$.*
4. *For every measurable set F, $x \to P^x(F)$ is universally measurable.*
5. *For each $x \in \mathcal{X}$, $P^x(X_{t+s} \in B | \mathcal{F}_s) = p_t(X_s, B)$ P^x-a.s. In particular, $p_t(x, B) = P^x(X_t \in B)$.*

Also assume we have a second collection of probability measures $\widetilde{P}^x$, $x \in \mathcal{X}$, on $(\Omega, \mathcal{F})$ that satisfy:

1. *$(X_0)_* \widetilde{P}^x = \delta_x$ for each $x \in \mathcal{X}$.*
2. *For every measurable set F, $x \to \widetilde{P}^x(F)$ is universally measurable.*

Note that we are *not* assuming X_t is a Markov processes for the $\widetilde{P}^x$.

One of the models is thought of as the base model, and the other as some alternative (or approximate) model, but which is which can vary with the application. From a mathematical perspective, the primary factors distinguishing P^x and $\widetilde{P}^x$ are:

1. Our methods require information on the spectrum of the generator of p_t.
2. $(X_t, \widetilde{P}^x)$ are not required to be Markovian.

P^x and $\widetilde{P}^x$ should be chosen with these points in mind.

Definition 2 Given initial distributions μ and $\widetilde{\mu}$ on $\mathcal{X}$, we also define the probability measures

$$P^\mu(\cdot) = \int P^x(\cdot) \mu(dx), \quad \widetilde{P}^{\widetilde{\mu}}(\cdot) = \int \widetilde{P}^x(\cdot) \widetilde{\mu}(dx). \tag{27}$$

Note that Assumption 1 implies that X_t is a Markov process for the space $(\Omega, \mathcal{F}_t, P^\mu)$ with initial distribution μ and time homogeneous transition probabilities p_t. Again, we make no such assumption regarding $\widetilde{P}^{\widetilde{\mu}}$.

We will also need the finite time restrictions, which can be though of as the distributions on path space up to some $T > 0$:

$$P_T^x \equiv P^x|_{\mathcal{F}_T}, \quad \widetilde{P}_T^x \equiv \widetilde{P}^x|_{\mathcal{F}_T}, \tag{28}$$

and similarly for P_T^μ and $\widetilde{P}_T^{\widetilde\mu}$. Finally, we let E^μ denote the expected value with respect to P^μ and similarly for $\widetilde{E}^{\widetilde\mu}$.

Now fix a bounded measurable $f : \mathcal{X} \to \mathbb{R}$ (the boundedness assumption will be relaxed later on) and an invariant measure μ^* for p_t. As mentioned in the introduction, there are many classical techniques for studying convergence of the ergodic averages of f under P^μ to the average in the invariant measure, $\mu^*[f]$. Therefore, in this paper we consider the much less-studied problem of bounding the bias when the finite-time averages are computed by sampling from the alternative distribution:

$$\textbf{Bound:} \qquad \widetilde{E}^{\widetilde\mu}\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \mu^*[f]. \tag{29}$$

2.4 UQ Bounds via the Feynman-Kac Semigroup

Due to our interest in the problem Eq. (29), we start the P-process in the invariant distribution μ^*, while the $\widetilde{P}$-process is started in an arbitrary distribution $\widetilde\mu$.

Given a bounded measurable function f on $\mathcal{X}$ and $T > 0$, define the bounded and $\mathcal{F}_T$-measurable function

$$f_T = \int_0^T f(X_t)dt. \tag{30}$$

Applying the Gibbs information inequality, Eq. (17), to f_T, $\widetilde{P}_T^{\widetilde\mu}$, $P_T^{\mu^*}$ and dividing by T yields:

Theorem 3

$$\pm\left(\widetilde{E}^{\widetilde\mu}\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \mu^*[f]\right) \le \varXi^\pm\left(\frac{1}{T}\varLambda_{P_T^{\mu^*}}^{\widehat{f}_T}, \frac{1}{T}R(\widetilde{P}_T^{\widetilde\mu}||P_T^{\mu^*})\right), \tag{31}$$

where

$$\mu^*[f] \equiv \int f d\mu^*, \quad \widehat{f}_T \equiv \int_0^T f(X_t) - \mu^*[f]dt. \tag{32}$$

Remark 3 Recall the definition

$$\varXi^\pm(\varLambda, \eta) = \inf_{c>0}\left\{\frac{\varLambda(\pm c) + \eta}{c}\right\}. \tag{33}$$

All of the UQ bounds we obtain will be of the form

$$\pm\left(\widetilde{E}^{\widetilde\mu}\left[\frac{1}{T}\int_0^T f(X_t)dt\right] - \mu^*[f]\right) \le \varXi^\pm(\varLambda, \eta) \tag{34}$$

for some $\varLambda : \mathbb{R} \to [0, \infty]$ and $\eta > 0$; we will refer back to these equations often.

To produce a more explicit bound from Eq. (31), one needs to bound the cumulant generating function as well as the relative entropy. The latter will be addressed in Section 6. As for the former, observe that the cumulant generating function can be written

$$\Lambda_{P_T^{\mu^*}}^{\widehat{f}_T}(\pm c) = \log\left(\int E^x\left[\exp\left(\pm c\int_0^T f(X_t) - \mu^*[f]dt\right)\right]\mu^*(dx)\right). \quad (35)$$

Eq. (35) is related to the Feynman-Kac semigroup on $L^2(\mu^*)$ with potential V:

$$\mathcal{P}_t^V[g](x) = E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right]. \quad (36)$$

More specifically,

$$\Lambda_{P_T^{\mu^*}}^{\widehat{f}_T}(\pm c) \leq \log\left(\|\mathcal{P}_T^{V_{\pm c}}[1]\|_{L^2(\mu^*)}\right), \quad (37)$$

$$V_{\pm c}(x) \equiv \pm\, c\left(f(x) - \mu^*[f]\right), \quad (38)$$

and so we obtain:

Lemma 3 *Under Assumption 1, for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$, Eq. (34) holds with*

$$\Lambda(\pm c) = \frac{1}{T}\log\left(\|\mathcal{P}_T^{V_{\pm c}}[1]\|_{L^2(\mu^*)}\right), \quad \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu^*}). \quad (39)$$

In the following two sections, we discuss how functional inequalities can be used to obtain more explicit bounds on the norm of the Feynman-Kac semigroup.

3 Bounding the Feynman-Kac Semigroup

The Lumer-Phillips theorem (a variant of the Hille-Yosida theorem) is our tool of choice for bounding the Feynman-Kac semigroup; see Chapter IX, p. 250 in [20]. This is the same strategy used in [11,13,15] to obtain concentration inequalities.

First we state some of the basic properties of the Feynman-Kac semigroup, adapted from [11,13].

Theorem 4 *Let $V : \mathcal{X} \to \mathbb{R}$ be bounded and measurable and μ^* be an invariant probability measure for p_t.*

For $t \geq 0$ define $\mathcal{P}_t^V : L^2(\mu^*) \to L^2(\mu^*)$,

$$\mathcal{P}_t^V[g](x) = E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right]. \quad (40)$$

These are bounded linear operators and form a strongly continuous semigroup.

If $(A, D(A))$ denotes the generator of $\mathcal{P}_t \equiv \mathcal{P}_t^0$ on $L^2(\mu^*)$ then the generator of $\mathcal{P}_t^V$ on $L^2(\mu^*)$ is $(A + V, D(A))$.

To bound the norm of the Feynman-Kac semigroup, we use the following Hilbert space version of the Lumer-Phillips theorem:

Theorem 5 *Let H be a Hilbert space and $Q(t)$ a strongly continuous semigroup on H with generator $(A, D(A))$. Suppose that there is an $\alpha \in \mathbb{R}$ such that*

$$\operatorname{Re}(\langle Ax, x \rangle) \leq \alpha \tag{41}$$

for all $x \in D(A)$ with $\|x\| = 1$. Then $\|Q(t)\| \leq e^{\alpha t}$ for all $t \geq 0$.

Theorems 4 and 5 together yield a bound on the Feynman-Kac semigroup, in terms of the generator; this result, and generalizations, were proven in [11].

Corollary 1 *Let $V : \mathcal{X} \to \mathbb{R}$ be bounded and measurable, and for $t \geq 0$ consider the Feynman-Kac semigroup $\mathcal{P}_t^V : L^2(\mu^*) \to L^2(\mu^*)$ with generator $(A + V, D(A))$.*

Define

$$\kappa(V) = \sup \left\{ \operatorname{Re}(\langle (A+V)[g], g \rangle) : g \in D(A), \|g\|_{L^2(\mu^*)} = 1 \right\} \tag{42}$$

$$= \sup \left\{ \langle A[g], g \rangle + \int V|g|^2 d\mu^* : g \in D(A, \mathbb{R}), \|g\|_{L^2(\mu^*)} = 1 \right\}, \tag{43}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2(\mu^*)$ and $D(A, \mathbb{R})$ denotes the real-valued functions in the domain of A.

Then the operator norm satisfies the bound

$$\|\mathcal{P}_t^V\| \leq e^{t\kappa(V)} \tag{44}$$

for all $t \geq 0$.

Combining Eq. (44) with Eq. (37) and Eq. (31), we obtain:

Corollary 2 *Under Assumption 1, for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$, the UQ bound Eq. (34) holds with*

$$\Lambda(\pm c) = \kappa(V_{\pm c}), \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu^*}). \tag{45}$$

From Eq. (34) we see that functional inequalities, by which we mean bounds on the generator A that lead to bounds on $\kappa(V_{\pm c})$, can be used to produce UQ bounds. Also, note that the only remaining T-dependence is in the relative entropy term, $R(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu^*})/T$. This will often have a finite limit (the relative entropy rate) as $T \to \infty$. Hence Corollary 2 shows that one can expect UQ bounds that are well behaved as $T \to \infty$.

Remark 4 Corollary 1 is stated for bounded V, but it can be extended to certain unbounded V under the additional assumption that the symmetrized Dirichlet form is closable; see Theorem 1 in [11]. However, as noted in Corollary 3 in this same reference (and outlined in Theorem 11 below), that assumption can be avoided in the presence of functional inequalities by working with bounded V and then taking limits; this is the strategy we employ here.

4 UQ Bounds From Functional Inequalities

In this section, we explore the consequences of several important classes of functional inequalities: Poincaré, log-Sobolev, and Liapunov functions. Discussion of F-Sobolov inequalities, a generalization of the classical log-Sobolev case, can be found in Appendix B.

4.1 Poincaré Inequality

First we consider the case where the generator satisfies a Poincaré inequality with constant $\alpha > 0$, meaning:

$$\mathrm{Var}_{\mu^*}(g) \leq -\alpha \langle A[g], g \rangle \tag{46}$$

for all $g \in D(A, \mathbb{R})$. This can equivalently be written

$$\mathrm{Re}(\langle A[g], g \rangle) \leq -\alpha^{-1} \|P^\perp g\|^2 \tag{47}$$

for all $g \in D(A)$, where $P^\perp$ is the orthogonal projector onto $1^\perp$.

In the presence of a Poincaré inequality, Corollary 1 is most useful when combined with the following perturbation result. A version of this result is contained in [11], but we present it here in a slightly more general form. The proof is given in Appendix A.

Lemma 4 *Let H be a Hilbert space, $A : D(A) \subset H \to H$ a linear operator and $B : H \to H$ a bounded self-adjoint operator. Suppose there exists $D > 0$ and $x_0 \in H$ with $\|x_0\| = 1$ such that*

$$\langle Bx_0, x_0 \rangle = 0 \quad and \quad \mathrm{Re}(\langle Ax, x \rangle) \leq -D\|P^\perp x\|^2 \tag{48}$$

for all $x \in D(A)$, where $P^\perp$ is the orthogonal projector onto $x_0^\perp$.

Define

$$B^+ \equiv \max \left\{ \sup_{\|y\|=1} \langle By, y \rangle, 0 \right\}. \tag{49}$$

Then for any $0 \leq c < D/B^+$ we have

$$\sup_{x \in D(A), \|x\|=1} \mathrm{Re}(\langle (A + cB)x, x \rangle) \leq \frac{c^2 \|Bx_0\|^2}{D - cB^+}. \tag{50}$$

The multiplication operator by $V_{\pm 1}$ is a bounded self-adjoint operator and $\langle V_{\pm 1} 1, 1 \rangle = 0$. Therefore Lemma 4 implies:

Lemma 5 *For all $0 \leq c < 1/(\alpha \|(f - \mu^*[f])^\pm\|_\infty)$ we have*

$$\kappa(V_{\pm c}) \leq \frac{\alpha \mathrm{Var}_{\mu^*}[f] c^2}{1 - \alpha \|(f - \mu^*[f])^\pm\|_\infty c}. \tag{51}$$

Thus we have shown the following UQ bound:

Theorem 6 *Under Assumption 1, if A satisfies the Poincaré inequality, Eq. (46), then for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$ the bounds Eq. (34) and Eq. (25) hold with*

$$M^{\pm} = \alpha \|(f - \mu^*[f])^{\pm}\|_{\infty}, \ \ \sigma^2 = 2\alpha \operatorname{Var}_{\mu^*}[f], \ \ \eta = \frac{1}{T} R(\widetilde{P}_T^{\tilde{\mu}} \| P_T^{\mu^*}). \tag{52}$$

4.2 Poincaré Inequality for Reversible Processes

When the combination of μ^* and p_t are reversible, i.e. the generator A is self-adjoint on $L^2(\mu^*)$, and if a Poincaré inequality, Eq. (46), also holds with constant $\alpha > 0$ then one can obtain a UQ bound in terms of the asymptotic variance, rather than the variance.

First, define the Poisson operator

$$L : f \to \int_0^{\infty} \mathcal{P}_t[f] dt, \tag{53}$$

a bounded linear operator on $L_0^2(\mu^*) \equiv \{f \in L^2(\mu^*) : \mu^*[f] = 0\}$ with norm bound $\|L\| \leq \alpha$. The asymptotic variance of $f \in L^2(\mu^*, \mathbb{R})$ is defined by

$$\sigma^2(f) \equiv \langle 2L[f - \mu^*[f]], f - \mu^*[f] \rangle = 2 \int_0^{\infty} \left(\int \mathcal{P}_t[f] f d\mu^* - (\mu^*[f])^2 \right) dt. \tag{54}$$

Note that $0 \leq \sigma^2(f) \leq 2\alpha \operatorname{Var}_{\mu^*}[f]$.

Using these objects, one can obtain the following Bernstein-type bound. A simple proof appears below Remark 2.3 in [15]; we outline the essential ideas below. See [12] and [14] for similar earlier results.

Lemma 6 *For all $0 < c < 1/(\alpha \|(f - \mu^*[f])^{\pm}\|_{\infty})$ we have*

$$\kappa(V_{\pm c}) \leq \frac{\sigma^2(f)c^2}{2(1 - \alpha \|(f - \mu^*[f])^{\pm}\|_{\infty} c)}. \tag{55}$$

Proof The cases where $\sigma^2(f) = 0$ or one of $\|(f - \mu^*[f])^{\pm}\|_{\infty} = 0$ are trivial, so suppose not. Using the self-adjoint functional calculus, one can see that L inverts A on $D(A) \cap L_0^2(\mu^*)$ and

$$\left| \int f g d\mu^* \right| \leq \left(\int -A[g] g d\mu^* \right)^{1/2} \left(\int -L[f] f d\mu^* \right)^{1/2} \tag{56}$$

for all real-valued $f \in L_0^2(\mu^*)$, $g \in D(A, \mathbb{R})$.

Hence, for any $g \in D(A, \mathbb{R})$ with $\|g\|_{L^2(\mu^*)} = 1$ and bounded V:

$$\int V g^2 d\mu^* = \int V(g - \mu^*[g])^2 d\mu^* + 2\mu^*[g] \int (V - \mu^*[V]) g d\mu^* + \mu^*[V] \mu^*[g]^2$$

$$\leq c \|V^+\|_{\infty} \operatorname{Var}_{\mu^*}(g) + \sqrt{2\sigma^2(V)} \sqrt{\langle -A[g], g \rangle} + \mu^*[V]. \tag{57}$$

Using the Poincaré inequality and solving for $\langle -A[g], g \rangle$ gives

$$\langle -A[g], g \rangle \geq h \left(\int (V - \mu^*[V]) g^2 d\mu^* \right), \tag{58}$$

$$h(r) \equiv 1_{r \geq 0} \frac{\sigma^2(V)}{2(M^\pm)^2} \left(\left(1 + \frac{2M^\pm}{\sigma^2(V)} r \right)^{1/2} - 1 \right)^2, \ \ M^\pm \equiv \alpha \|V^+\|_\infty.$$

Letting $V = V_{\pm 1}$ and substituting Eq. (58) into the expression for κ, Eq. (43), results in

$$\kappa(V_{\pm c}) \leq \sup_{r \in \mathbb{R}} \{ cr - h(r) \}. \tag{59}$$

Eq. (55) then follows from solving the optimization problem.   □

The Bernstein-type bound, Eq. (55), implies a UQ bound similar to Theorem 6:

Theorem 7 *Under Assumption 1, if the generator satisfies the Poincaré inequality Eq. (46) and is self-adjoint on $L^2(\mu^*)$ then for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$ the bounds Eq. (34) and Eq. (25) hold with*

$$M^\pm = \alpha \|(f - \mu^*[f])^\pm\|_\infty, \ \ \sigma^2 = \sigma^2(f), \ \ \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu^*}). \tag{60}$$

Other variations can be derived using a Liapunov function. First we need a couple of definitions, taken from [15]. Also, see this reference for further Liapunov function results that could likely be adapted to produce UQ bounds.

Definition 3 A measurable function $G : \mathcal{X} \to \mathbb{R}$ is in the μ^*-extended domain of the generator, $D_{e,\mu^*}(A)$, if there is some measurable $g : \mathcal{X} \to \mathbb{R}$ such that $\int_0^t |g|(X_s) ds < \infty$ P^{μ^*}-a.s. and one P^{μ^*}-version of

$$M_t(G) \equiv G(X_t) - G(X_0) - \int_0^t g(X_s) ds \tag{61}$$

is a local P^{μ^*}-martingale.

$U \in D_{e,\mu^*}(A)$ is called a Liapunov function if there exists a measurable $\phi : \mathcal{X} \to (0, \infty)$ and $b > 0$ such that

$$-\frac{A[U]}{U} \geq \phi - b \ \ \mu^*\text{-a.s.} \tag{62}$$

As shown in [15], given a Liapunov function one can derive a bound on $\kappa(V_{\pm c})$; our method then produces a corresponding UQ bound:

Theorem 8 *In addition to Assumption 1, assume the generator, A, is self-adjoint on $L^2(\mu^*)$, satisfies the Poincaré inequality Eq. (46), and we have a Liapunov function U with $-A[U]/U \geq \phi - b$.*

Given an observable $f \in L^2(\mu^*, \mathbb{R})$ with $\|(f - \mu^*[f])^\pm/\phi\|_\infty < \infty$, we have the UQ bounds Eq. (34) and Eq. (25), where

$$M^\pm = (1 + \alpha b)\|(f - \mu^*[f])^\pm/\phi\|_\infty, \ \ \sigma^2 = \sigma^2(f), \ \ \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu^*}). \quad (63)$$

Proof First let V be a bounded measurable function. This part of the proof proceeds similarly to that of Lemma 6, but rather than taking the supremum of V^+ in Eq. (57), one instead uses Eq. (62) to compute the following bound, where $g \in D(A, \mathbb{R})$ with $\|g\|_{L^2(\mu^*)} = 1$:

$$\int Vg^2 d\mu^* \leq \mu^*[V] + \sqrt{2\sigma^2(V)}\sqrt{\langle -A[g], g \rangle} \quad (64)$$

$$+ \|V^+/\phi\|_\infty \int \left(-\frac{A[U]}{U} + b\right)(g - \mu^*[g])^2 d\mu^*.$$

Next, use the bound found in Lemma 5.6 in [14],

$$\int -\frac{A[U]}{U}(g - \mu^*[g])^2 d\mu^* \leq \langle -A[g], g \rangle, \quad (65)$$

and proceed as in Lemma 6 to obtain

$$\kappa(\pm cV) \leq \pm c\mu^*[V] + \frac{\sigma^2(V)c^2}{2(1 - (1 + \alpha b)\|V^\pm/\phi\|_\infty c)} \quad (66)$$

for all $0 < c < 1/((1 + \alpha b)\|V^\pm/\phi\|_\infty)$. If f is bounded then letting $V = f - \mu^*[f]$ and using Corollary 2 and Lemma 1 gives the claimed UQ bound.

For general $f \in L^2(\mu^*, \mathbb{R})$ with $\|(f - \mu^*[f])^\pm/\phi\|_\infty < \infty$, we employ a similar method to Corollary 3 in [11]: Define $V = f - \mu^*[f]$ and $V_n = V1_{|V| < n}$. Applying the above result to V_n and then using Fatou's Lemma and L^2-continuity of the asymptotic variance gives

$$\frac{1}{T}\Lambda_{P_T^{\mu^*}}^{\widehat{f}_T}(\pm c) \leq \frac{1}{T}\log\left(\|\mathcal{P}_T^{V_{\pm c}}[1]\|\right) \leq \liminf_{n\to\infty}\frac{1}{T}\log\left(\|\mathcal{P}_T^{\pm cV_n}[1]\|\right) \quad (67)$$

$$\leq \liminf_{n\to\infty}\left(\pm c\mu^*[V_n] + \frac{\sigma^2(V_n)c^2}{2(1 - (1 + \alpha b)\|(f - \mu^*[f])^\pm/\phi\|_\infty c)}\right)$$

$$= \frac{\sigma^2(f)c^2}{2(1 - (1 + \alpha b)\|(f - \mu^*[f])^\pm/\phi\|_\infty c)}.$$

Having extended the bound on the cumulant generating function to such f, the claimed UQ bound follows from Theorem 1. $\square$

4.3 Poincaré Inequality Examples

The study of Poincaré inequalities has a long history which we do not attempt to recount here; rather, we simply present several examples.

4.3.1 Continuous-Time Markov Chains

If we consider a continuous-time Markov chain on a finite state-space $\mathcal{X}$ with jump-rates λ_i and transition probabilities $a_{i,j}$ then the generator of the continuous-time transition semigroup is $Q_{i,j} = \lambda_i(a_{i,j} - \delta_{i,j})$.

A Poincaré inequality for Q holds with constant $\alpha > 0$ and in the invariant measure μ^* if and only if the eigenvalues of the symmetric part of Q consist of 0, with eigenspace spanned by 1, and all other eigenvalues are bounded above by $-\alpha^{-1}$. Note that the symmetric part of Q is defined relative to $L^2(\mu^*)$ i.e. $Q_{i,j}^* = (\mu^*)_j Q_{j,i}/(\mu^*)_i$.

4.3.2 Diffusions

Diffusion processes in a potential that grows sufficiently fast at infinity provide another important class of examples that satisfy a Poincaré inequality. Specifically, let V be C^2 and bounded below such that $\mu^*(dx) = e^{-V(x)}dx$ is a probability measure, and consider the diffusion process with generator $A = \Delta - \nabla V \cdot \nabla$.

A proof of the following simple suffcent conditions for such systems to satisfy a Poincaré inequality can be found in [21].

Suppose V satisfies either of the following:

1. There exists $a > 0$, $R \geq 0$ such that for $|x| \geq R$,

$$x \cdot \nabla V(x) \geq a|x|. \tag{68}$$

2. There exists $a \in (0,1)$, $b > 0$, and $R \geq 0$ such that for $|x| \geq R$,

$$a|\nabla V(x)|^2 - \Delta V(x) \geq b. \tag{69}$$

Then a Poincaré inequality holds on $L^2(\mu^*)$; see [21] for details on the form of this constant. Other conditions can be found in [22].

4.3.3 Poincaré Inequality from Exponential Convergence

In the self-adjoint case, a Poincaré inequality for A (or equivalently, a spectral gap) can be obtained from exponential convergence bounds in other norms. First, a key lemma involving the L^2 norm:

Lemma 7 *Suppose $(A, D(A))$ is self-adjoint, $D \subset L^2(\mu^*)$ has dense span, and there exists $\alpha > 0$ such that the following holds:*
For every $f \in D$ there exists $C_f \geq 0$ and $t_n^f \geq 0$ with $t_n^f \to \infty$, such that

$$\|\mathcal{P}_{t_n^f}[f] - \mu^*[f]\|_2 \leq C_f e^{-t_n^f/\alpha} \tag{70}$$

for all n.

Then a Poincaré inequality, Eq. (46), holds with constant α.

See, for example, Lemma 3.8 in [23].

As an example of this Lemma's utility, the following result shows that a Poincaré inequality (with an explicit constant) can be deduced from exponential convergence in a pair of weighted norms.

Theorem 9 *Suppose $(A, D(A))$ is self-adjoint, and $W : \mathcal{X} \to [1, \infty)$ is measurable. Define the following norms on measurable functions $\phi : \mathcal{X} \to \mathbb{R}$ and signed measures π on $\mathcal{X}$:*

$$|\phi|_W = \sup_{x \in \mathcal{X}} \frac{|\phi(x)|}{W(x)}, \quad |\pi|_W = \int W d|\pi|. \tag{71}$$

Suppose we have $\lambda \geq 0, \rho \geq 0$ with at least one nonzero, and that for every bounded measurable $h : \mathcal{X} \to [0, \infty)$ with $\int h d\mu = 1$ there exists $t_n^h \geq 0$, $n \in \mathbb{Z}^+$, converging to ∞ and $C_h, D_h \in [0, \infty)$ such that for all n:

$$|\mathcal{P}_{t_n^h}[h] - 1|_W \leq D_h e^{-\rho t_n^h} \tag{72}$$

and the measure $d\nu = h d\mu^*$ satisfies

$$|\mathcal{P}_{t_n^h}^*[\nu] - \mu^*|_W \leq C_h e^{-\lambda t_n^h}. \tag{73}$$

Then A satisfies the Poincare inequality

$$\mathrm{Var}_{\mu^*}(g) \leq -\frac{2}{\lambda + \rho} \langle A[g], g \rangle \text{ for all } g \in D(A, \mathbb{R}). \tag{74}$$

Proof The proof is similar to that of Theorem 2.1 in [22]. The key is to take h as above, let $d\nu = h d\mu^*$, and use symmetry of $\mathcal{P}_t$ to compute

$$\|\mathcal{P}_t[h] - 1\|_2^2 = \int \frac{|\mathcal{P}_t[h] - 1|}{W} W |\mathcal{P}_t[h] - 1| d\mu^* \leq |\mathcal{P}_t[h] - 1|_W |\mathcal{P}_t^*[\nu] - \mu^*|_W. \tag{75}$$

$\square$

In [24, 25], it is shown how exponential convergence in the norms $|\cdot|_W$ can be deduced from existence of a Liapunov function. The following is a summary:

Theorem 10 (Harris' Theorem) *Let $T > 0$ and suppose we have a Liapunov function for $\mathcal{P}_T$, meaning a measurable $V : \mathcal{X} \to [0, \infty)$ for which*

$$\mathcal{P}_T[V](x) \leq \gamma V(x) + K \tag{76}$$

for some $K \geq 0$, $\gamma \in (0, 1)$ and all $x \in \mathcal{X}$.

Also assume there exists $R > 2K/(1 - \gamma)$, $\alpha \in (0, 1)$ such that one of the following conditions holds:

1. *For all $x, y \in \mathcal{X}$ with $V(x) + V(y) \leq R$ we have*

$$\|p_T(x, \cdot) - p_T(y, \cdot)\|_{TV} \leq 2(1 - \alpha). \tag{77}$$

2. *There exists a probability measure ν on $\mathcal{X}$ such that*

$$\inf_{x:V(x)<R} p_T(x,\cdot) \geq \alpha\nu. \tag{78}$$

Let $\alpha_0 \in (0,\alpha)$ and define $\gamma_0 = \gamma + 2K/R$, $\beta = \alpha_0/K$, $W = 1 + \beta V$, and $\xi = \max\{1 - (\alpha - \alpha_0), (2 + R\beta\gamma_0)/(2 + R\beta)\}$.

Under the above conditions, $\mathcal{P}_T$ has unique invariant measure μ^*, $|\mu^*|_W < \infty$, for any probability measure ν with $|\nu|_W < \infty$ there exists $C_\nu \in [0,\infty)$ such that

$$|\mathcal{P}_{nT}^*[\nu] - \mu^*|_W \leq C_\nu \xi^n, \tag{79}$$

and for any measurable ϕ with $|\phi|_W < \infty$ there exists $C_\phi \in [0,\infty)$ such that

$$|\mathcal{P}_{nT}[\phi] - \mu^*[\phi]|_W \leq C_\phi \xi^n. \tag{80}$$

In particular, Eq. (72) and Eq. (73) hold with $\rho = \lambda = T^{-1}\log(1/\xi)$.

The required Liapunov function bound on $\mathcal{P}_T$ often follows from a related bound on the generator. For instance, when A is the generator of a stochastic differential equation (SDE) on $\mathbb{R}^n$ and V is C^2, Dynkin's formula generally lets one transform a bound of the form

$$A[V](x) \leq -cV(x) + C \text{ for all } x \tag{81}$$

into

$$\mathcal{P}_T[V](x) \leq e^{-cT}V(x) + C(1 - e^{-cT})/c \text{ for all } x. \tag{82}$$

4.3.4 A simple Liapunov example: the $M/M/\infty$ queue.

Following [15], let us consider the (simple) example of a $M/M/\infty$ queuing system which has infinitely many servers, each of which with a service rate ρ and with an arrival rate λ. The state space is $\mathbb{N}$ and the generator is given by

$$A[f](n) = \lambda f(n+1) - (\lambda + \rho n)f(n) + \rho n f(n-1) \tag{83}$$

The invariant measure μ^* is a Poisson distribution with paramter λ/ρ. An explicit computation shows (see e.g. [26]) that $\mathrm{Var}_{\mu^*}(\mathcal{P}_t f) \leq e^{-2\rho t} \mathrm{Var}_{\mu^*}(f)$ and thus the Poincaré constant is $1/\rho$.

To construct a Liapunov function take $U(n) = \kappa^n$ with $\kappa > 1$ and then we have

$$-\frac{A[U]}{U}(n) = \rho n(1 - \kappa^{-1}) - \lambda(\kappa - 1), \tag{84}$$

and we can apply Theorem 8 to any function f with $|f| \leq C(n + \delta)$ for some $\delta > 0$.

It is instructive to consider further the case of the mean number of customers in the queue, i.e., $f = n$ and $\widehat{f} = f - \mu^*[f] = n - \lambda/\rho$. From Eq. (83) we obtain

$$(A + \rho(1 - \kappa^{-1})\widehat{f})[U](n) = \lambda\frac{(\kappa - 1)^2}{\kappa}U(n) \tag{85}$$

and thus U is an eigenvector for $A + \rho(1 - \kappa^{-1})\widehat{f}$ with eigenvalue $\lambda\frac{(\kappa-1)^2}{\kappa}$. By Perron-Frobenius theorem and Rayleigh's principle we obtain that

$$\Lambda(c) \equiv \lim_{T \to \infty} \Lambda_{P_T^{\mu^*}}^{\widehat{f_T}}(c) \tag{86}$$

is the maximal eigenvalue of $A + c\widehat{f}$ and thus $\Lambda\left(\rho(1-\kappa^{-1})\right) = \lambda\frac{(\kappa-1)^2}{\kappa}$ or equivalently $\Lambda(c) = \frac{\lambda c^2}{\rho^2(1-c\rho^{-1})}$. Since $A\widehat{f}(n) = \lambda - \rho n$ we can solve the Poisson equation: $(-A)^{-1}\widehat{f} = \widehat{f}/\rho$ and thus the asymptotic variance is $\sigma^2(f) = 2\langle(-A)^{-1}\widehat{f}, \widehat{f}\rangle = 2\rho\,\mathrm{Var}_{\mu^*}[f] = \lambda/\rho^2$. As a consequence we have

$$\Lambda(c) = \frac{\sigma^2(f)}{2(1 - c\rho^{-1})} \tag{87}$$

which shows that Bernstein bounds can be sharp in the context of Markov processes, contrary to the IID setting.

4.4 log-Sobolev Inequalities

Next consider the log-Sobolev inequality with constant $\beta > 0$:

$$\int g^2 \log(g^2)d\mu^* \leq -\beta \int A[g]gd\mu^* \tag{88}$$

for all $g \in D(A, \mathbb{R})$ with $\|g\|_{L^2(\mu^*)} = 1$.

We will employ the following generalization of the Feynman-Kac semigroup for (possibly) unbounded potentials. The subsequent theorem was shown in Corollary 4 in [11]. For completeness purposes, we outline the proof.

Theorem 11 *Let A be the generator of $\mathcal{P}_t$ and μ^* be an invariant measure for the adjoint semigroup, $\beta > 0$, and assume the* log-*Sobolev inequality, Eq. (88), holds for μ^* with constant β.*

Finally, suppose that $V \in L^1(\mu^*)$ with $\int e^{\beta V} d\mu^* < \infty$. Then $\mathcal{P}_t^V : L^2(\mu^*) \to L^2(\mu^*)$, defined by

$$\mathcal{P}_t^V[g](x) = E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right], \tag{89}$$

are well-defined linear operators and the operator norm satisfies the bound

$$\|\mathcal{P}_t^V\| \leq \left(\int e^{\beta V} d\mu^*\right)^{t/\beta}. \tag{90}$$

Proof First assume V is bounded. Eq. (44) gives $\|\mathcal{P}_t^V\| \leq e^{t\kappa(V)}$. Applying the log-Sobolev inequality together with the Gibbs Variational principle, Eq. (16), we obtain

$$\kappa(V) \leq \beta^{-1}\sup\left\{-\int g^2\log(g^2)d\mu^* + \int \beta V|g|^2 d\mu^* : \|g\|_{L^2(\mu^*)} = 1\right\} \tag{91}$$

$$= \beta^{-1}\sup_{d\nu = g^2 d\mu^* : \|g\|_2 = 1}\{E_\nu[\beta V] - R(\nu\|\mu^*)\} = \beta^{-1}\log\left(\int \exp(\beta V)d\mu^*\right),$$

which proves the claim.

The case of unbounded V satisfying the assumptions of the theorem is obtained by letting $V_n = V1_{|V|\leq n}$, and then using Fatou's lemma, the result for bounded V, and dominated convergence to compute

$$\|\mathcal{P}_t^V\| \leq \liminf_{n\to\infty} \|\mathcal{P}_t^{V_n}\| \leq \liminf_{n\to\infty} \left(\int e^{\beta V_n} d\mu^*\right)^{t/\beta} = \left(\int e^{\beta V} d\mu^*\right)^{t/\beta}.$$

□

Using Theorem 11, a UQ bound of the form Eq. (34) can be derived that covers a class of unbounded observables:

Theorem 12 *In addition to Assumption 1, assume the* log-*Sobolev inequality, Eq. (88), holds and we have an observable $f \in L^1(\mu^*, \mathbb{R})$ and $c_- < 0 < c_+$ such that for all $c \in (c_-, c_+)$:*

$$\int \exp\left(\beta V_{\pm c}\right) d\mu^* < \infty. \tag{92}$$

Then a UQ bound of the form Eq. (34) holds with

$$\Lambda(c) = \begin{cases} \frac{1}{\beta} \log\left(\int e^{\beta V_c} d\mu^*\right) & \text{if } c \in (c_-, c_+) \\ +\infty & \text{otherwise} \end{cases}. \tag{93}$$

In addition, the asymptotic result Eq. (26) holds with

$$\Lambda''(0) = \beta \operatorname{Var}_{\mu^*}[f], \quad \eta = \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} \| P_T^{\mu^*}). \tag{94}$$

Proof The bound Eq. (90) implies $E^{\mu^*}[\exp(cf_T)] < \infty$ for $c \in (c_-, c_+)$, hence $f_T \in \mathcal{E}(P_T^{\mu^*})$ and the Gibbs information inequality, Eq. (17), applies. As in Eq. (37), the cumulant generating function can be bounded using the Feynman-Kac semigroup bound, Eq. (90). Combining this with Eq. (17) yields a bound of the form Eq. (34), with Λ as defined in Eq. (93).   □

The ideas in this section can be extended to F-Sobolev inequalities; see Appendix B.

4.4.1 Example: Diffusions

Again consider the diffusion example with generator $A = \Delta - \nabla V \cdot \nabla$ and invariant measure $\mu^*(dx) = e^{-V(x)} dx$ discussed in Section 4.3.2. First, it is useful to note that a log-Sobolev inequality with constant β implies a Poincaré inequality with constant $\alpha = \beta/2$ [27].

In [28], the following sufficient condition for a log-Sobolev inequality was obtained:

Suppose A satisfies a Poincaré inequality with constant α and that

$$-C \equiv \inf_x \left\{ \frac{1}{4}|\nabla V(x)|^2 - \frac{1}{2}\Delta V(x) - \pi e^2 V(x) \right\} > -\infty. \tag{95}$$

Then A satisfies a log-Sobolev inequality with constant

$$\beta = 3\alpha + \frac{1}{(1 + \alpha|C|)\pi e^2}. \tag{96}$$

Hypotheses guaranteeing a Poincaré inequality were discussed in Section 4.3.2.

Additionally, if the Hessian of V is bounded below,

$$D^2V(x) \geq 2\beta^{-1}I, \;\; \beta > 0, \tag{97}$$

then a log-Sobolev inequality holds with constant β [29]. The UQ bound corresponding to the associated Poincaré inequality with constant $\alpha = \beta/2$ was given in the introduction, Eq. (11).

5 Functional Inequalities and UQ for Discrete-Time Markov Processes

In this section we show how the above framework can be applied to obtain UQ bounds for invariant measures of discrete-time Markov processes.

Again, let $\mathcal{X}$ be a Polish space, and suppose we have time-homogeneous one-step transition probabilities $p(x, dy)$ and $\widetilde{p}(x, dy)$ on $\mathcal{X}$ with invariant measures μ^* and $\widetilde{\mu}^*$ respectively. Assume that $R(\widetilde{\mu}^*||\mu^*) < \infty$.

Define the bounded linear operator $\mathcal{P}$ on $L^2(\mu^*)$,

$$\mathcal{P}[f](x) \equiv \int f(y)p(x, dy), \tag{98}$$

and similarly for $\widetilde{\mathcal{P}}$ on $L^2(\widetilde{\mu}^*)$.

We obtain UQ bounds for expectations in μ^* and $\widetilde{\mu}^*$ by constructing continuous-time processes with these same invariant distributions. Specifically, in Appendix C (see Theorem C3) we obtain càdlàg Markov families $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \{P^x\}_{x \in \mathcal{X}})$ and $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, \{\widetilde{P}^x\}_{x \in \mathcal{X}})$, whose transition probabilities p_t and $\widetilde{p}_t$, respectively, (not to be confused with p and $\widetilde{p}$) satisfy the following:

1. μ^* is invariant for p_t for all $t \geq 0$, and similarly for $\widetilde{\mu}^*$ and $\widetilde{p}_t$ (see Theorem C4).

2. The continuous-time semigroup, $\mathcal{P}_t$, on $L^2(\mu^*)$ constructed from p_t is

$$\mathcal{P}_t = \exp(t(\mathcal{P} - I)). \tag{99}$$

   Specifically, $\mathcal{P}_t$ has bounded generator $A = \mathcal{P} - I$ (see Theorem C4). Note that we will also refer to A as the generator of the discrete-time Markov process.

3. The relative entropy of rate of the continuous-time process can be bounded by the relative entropy of the discrete-time process as follows:

$$R(\widetilde{P}_T^{\widetilde{\mu}^*}||P_T^{\mu^*}) \leq R(\widetilde{\mu}^*||\mu^*) + T \int R(\widetilde{p}(x, \cdot)||p(x, \cdot))\widetilde{\mu}^*(dx) \tag{100}$$

   for all $T > 0$ (see Theorem C6 and Corollary C1).

In particular, Assumption 1 holds for P^x and $\widetilde{P}^x$. If the generator $\mathcal{P} - I$ satisfies any of the functional inequalities covered in Section 3 then the general results therein imply UQ bounds for expectations in the invariant measures μ^* and $\widetilde{\mu}^*$, with Eq. (100) providing a bound on the relative entropy rate.

Remark 5 Note that here, we must take $\widetilde{\mu} = \widetilde{\mu}^*$ for the bounds to apply to the original discrete-time process, otherwise one obtains UQ bounds for ergodic averages of $f(X_t)$ under the auxiliary continuous-time Markov family.

For example, a Poincaré inequality for the generator $\mathcal{P} - I$,

$$\mathrm{Re}(\langle (\mathcal{P} - I)g, g\rangle) \leq -\alpha^{-1}\|P^\perp g\|^2_{L^2(\mu^*)}, \ \ g \in L^2(\mu^*), \ \alpha > 0, \tag{101}$$

implies that for any bounded measurable $f : \mathcal{X} \to \mathbb{R}$, we have the Bernstein-type UQ bound:

$$\pm (\widetilde{\mu}^*[f] - \mu^*[f]) \leq \sqrt{2\sigma^2\eta} + M^\pm\eta, \tag{102}$$
$$\sigma^2 = 2\alpha \, \mathrm{Var}_{\mu^*}[f], \ \ M^\pm = \alpha\|(f - \mu^*[f])^\pm\|_\infty,$$
$$\eta = \int R(\widetilde{p}(x, \cdot)\|p(x, \cdot))\widetilde{\mu}^*(dx).$$

This follows from Theorem 6, after taking $T \to \infty$.

We illustrate these discrete-time UQ bounds with a pair of examples:

5.1 Example: Random Walk on a Hypercube

Consider the symmetric random walk on the d-dimensional hypercube $\mathcal{X} = \{-1, 1\}^d$ i.e. the transition probabilities are defined by uniformly randomly selecting a coordinate, $i \in \{1, ..., d\}$, and then independently and uniformly selecting the sign, 1 or -1, with which to update the selected component.

The uniform measure, μ^*, on $\mathcal{X}$ is invariant and the process is reversible on $(\mathcal{X}, \mu^*)$. The eigenvalues and eigenvectors of the transition matrix can be found explicitly; see example 12.15 in [30]. In particular, the second largest eigenvalue is $\lambda_2 = 1 - 1/d$, hence we obtain the following Poincaré inequality:

$$\mathrm{Re}(\langle (\mathcal{P} - I)g, g\rangle) \leq -\frac{1}{d}\|P^\perp g\|^2_{L^2(\mu^*)}, \ \ g \in L^2(\mu^*). \tag{103}$$

Therefore, assuming $R(\widetilde{\mu}^*\|\mu^*) < \infty$, we obtain the UQ bound Eq. (102) with $\alpha = d$.

5.2 Example: Exclusion Chain

Derivation of functional inequalities for many discrete-time Markov processes can be found in [31]. Here we investigate the resulting UQ bounds for one of these examples; see Section 4.6 in the above reference and also [32] for further details and proofs regarding this example.

Let (V, E) be a symmetric, connected graph with n vertices. Let $d(x)$ be the degree of a vertex $x \in V$ and $d_0 = \max_x d(x)$. Fix $r \leq n$. The r-exclusion process is a Markov chain with state space being the set of cardinality r subsets of V. Informally stated, the transition probabilities are defined as follows: Given an r-subset A, pick an element $x \in A$ with probability proportional to its degree. Uniformly randomly pick a vertex y out of all those connected with x. If y is not in A then transition to the set $(A \setminus \{x\}) \cup \{y\}$. Otherwise, the chain remains at A.

For each $(x, y) \in V \times V$, fix a path $\gamma_{x,y}$ from x to y in the graph and let $|\gamma_{x,y}|$ be its length. Define

$$\Delta_0 = \max_{e_0 \in V} \left\{ \sum_{(x,y):e_0 \in \gamma_{x,y}} |\gamma_{x,y}| \right\}, \quad d_r = \max_{A \subset V: |A| = r} \left\{ \frac{1}{r} \sum_{a \in A} d(a) \right\}. \quad (104)$$

The generator of this Markov chain satisfies both a Poincaré inequality and a log-Sobolev inequality with respective constants being

$$\alpha = r d_r \Delta_0 / n, \quad \beta = 3 r d_r \Delta_0 \log(n) / n. \quad (105)$$

Then, assuming $R(\widetilde{\mu}^* || \mu^*) < \infty$, the above Poincaré inequality implies the UQ bound Eq. (102) with α as in Eq. (105), and the log-Sobolev inequality results in

$$\pm \left(\widetilde{\mu}^*[f] - \mu^*[f] \right) \leq \inf_{c > 0} \left\{ \frac{1}{c\beta} \log \left(\int \exp \left(\pm \beta c (f - \mu^*[f]) \right) d\mu^* \right) + \frac{\eta}{c} \right\}, \quad (106)$$

with β and η as in Eq. (105) and Eq. (102) respectively.

6 Bounding the Relative Entropy Rate

For any $\eta > 0$, the results derived in the previous sections provide UQ bounds over the class of all alternative models that satisfy a relative entropy bound of the form

$$H_T(\widetilde{P}^{\widetilde{\mu}} || P^{\mu^*}) \equiv \frac{1}{T} R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^{\mu^*}) \leq \eta. \quad (107)$$

In this section, we study in more detail the dependence of H_T on T and on the models $\widetilde{P}^{\widetilde{\mu}}$ and P^{μ^*}. Specifically, we derive upper bounds on H_T in various settings that can be substituted for H_T in the general UQ bound Eq. (34). Here, it will make little difference whether the initial distribution for the P-process is invariant or not, so we no longer make that assumption when deriving the relative entropy bounds; μ will denote an arbitrary initial distribution.

Deriving bounds on the relative entropy is a very application-specific problem. We will cover several examples in detail: continuous-time Markov chains, change of drift in SDEs, and numerical methods for SDEs with additive noise.

6.1 Example: Continuous-Time Markov Chains

Let $\mathcal{X}$ be a countable set, P^μ, $\widetilde{P}^{\widetilde{\mu}}$ be probability measures on $(\Omega, \mathcal{F})$ and $X_t : \Omega \to \mathcal{X}$ such that P^μ (resp. $\widetilde{P}^{\widetilde{\mu}}$) make $(\Omega, \mathcal{F}, X_t)$ a continuous-time Markov chain with transition probabilities $a(x, y)$ (resp. $\widetilde{a}(x, y)$), jump rates $\lambda(x)$ (resp. $\widetilde{\lambda}(x)$), and initial distribution μ (resp. $\widetilde{\mu}$). Let $\mathcal{F}_t$ be the natural filtration for X_t, X_n^J be the embedded jump chain with jump times J_n.

Suppose $\widetilde{\mu} \ll \mu$, $\forall x \forall y (a(x, y) = 0 \iff \widetilde{a}(x, y) = 0)$, and λ and $\widetilde{\lambda}$ are positive and bounded above. Then for any $T > 0$ we have $\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_T} \ll P^\mu|_{\mathcal{F}_T}$ and

$$R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_T} || P^\mu|_{\mathcal{F}_T}) \tag{108}$$

$$= R(\widetilde{\mu}||\mu) + \widetilde{E}^{\widetilde{\mu}}\left[\int_0^T \widetilde{F}(X_s)\widetilde{\lambda}(X_s)ds\right] - \widetilde{E}^{\widetilde{\mu}}\left[\int_0^T \widetilde{\lambda}(X_s) - \lambda(X_s)ds\right],$$

$$\widetilde{F}(x) \equiv \sum_{z \in \mathcal{X}} \widetilde{a}(x, z) \log\left(\frac{\widetilde{\lambda}(x)\widetilde{a}(x, z)}{\lambda(x)a(x, z)}\right).$$

To simplify further, if $\widetilde{\mu} = \widetilde{\mu}^*$ is an invariant measure then

$$R(\widetilde{P}^{\widetilde{\mu}^*}|_{\mathcal{F}_T} || P^\mu|_{\mathcal{F}_T}) = R(\widetilde{\mu}^*||\mu) \tag{109}$$

$$+ T\left(\sum_{x \in \mathcal{X}}\sum_{z \in \mathcal{X}} \widetilde{\mu}^*(x)\widetilde{\lambda}(x)\widetilde{a}(x, z) \log\left(\frac{\widetilde{\lambda}(x)\widetilde{a}(x, z)}{\lambda(x)a(x, z)}\right) - \sum_{x \in \mathcal{X}} \widetilde{\mu}^*(x)\left(\widetilde{\lambda}(x) - \lambda(x)\right)\right).$$

See the supplementary materials to [2] and Prop. 2.6 in App. 1 of [33] for details regarding these results.

6.2 Example: Change of Drift for SDEs

Next, consider the case where P^x and $\widetilde{P}^x$ are the distributions on $C([0, \infty), \mathbb{R}^n)$ of the solution flows X_t^x and $\widetilde{X}_t^x$ of a pair of SDEs. More precisely:

Assumption 2 *Assume:*

1. *X_t^x and $\widetilde{X}_t^x$ are weak solutions to the $\mathbb{R}^n$-valued SDEs, on filtered probability spaces satisfying the usual conditions [34]:*

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)dW_t, \ X_0^x = x, \tag{110}$$

$$d\widetilde{X}_t^x = \widetilde{b}(\widetilde{X}_t^x)dt + \sigma(\widetilde{X}_t^x)d\widetilde{W}_t, \ \widetilde{X}_0^x = x, \tag{111}$$

where W_t and $\widetilde{W}_t$ are (possibly different) m-dimensional Wiener processes. We let P and $\widetilde{P}$ denote the probability measures of the respective spaces where the SDEs are defined.
Here we think of $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ as the measurable drift and diffusion for the base process, and we assume the modified drift has the form $\widetilde{b} = b + \sigma\beta$ for some measurable $\beta : \mathbb{R}^n \to \mathbb{R}^n$.

2. X_t^x and $\widetilde{X}_t^x$ are jointly continuous in (t, x).
3. X_t^x satisfies the following flow property:
   For any bounded, measurable $G : C([0, \infty), \mathcal{X}) \to \mathbb{R}$, we have

$$E_P(G(X^x|_{t+[0,\infty)})|\mathcal{F}_t) = E_P\left[G\left(X^{(\cdot)}\right)\right] \circ X_t^x. \tag{112}$$

4. X_t^x and β satisfy the Novikov condition

$$E_P\left[\exp\left(\frac{1}{2}\int_0^T \|\beta(X_s^x)\|^2 ds\right)\right] < \infty \tag{113}$$

   for all $x \in \mathbb{R}^n$, $T > 0$.
5. For every $T > 0$, solutions to Eq. (111) satisfy uniqueness in law, up to time T.

Given this, we define $P^x = (X^x)_* P$ and $\widetilde{P}^x = (\widetilde{X}^x)_* \widetilde{P}$ i.e. the distributions on path space, with the Borel sigma algebra:

$$(\Omega, \mathcal{F}, \mathcal{F}_t) = (C([0, \infty), \mathbb{R}^n), \mathcal{B}(C([0, \infty), \mathbb{R}^n)), \sigma(\pi_s, s \leq t)), \tag{114}$$

where π_t is evaluation at time t. Finally, define $X_t \equiv \pi_t$. One can easily show that the above properties are sufficient to guarantee that Assumption 1 holds.

Remark 6 The existence of flows of solutions X_t^x and $\widetilde{X}_t^x$ that satisfy the above conditions is guaranteed, for example, if b and σ satisfy a linear growth bound

$$\|b(x)\|^2 + \|\sigma(x)\|^2 \leq K^2(1 + \|x\|^2), \tag{115}$$

and the following local Lipschitz bound:
For each n there exists K_n such that

$$\|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| \leq K_n\|x - y\| \tag{116}$$

on $\|x\|, \|y\| \leq n$, and if $\beta : \mathbb{R}^n \to \mathbb{R}^n$ is also bounded and locally Lipschitz.

Fixing $T > 0$, Girsanov's theorem allows one to bound the relative entropy, $R(\widetilde{P}_T^x || P_T^x)$, that appears in the UQ bound Eq. (34). See the supplementary materials to [2] for more details:

Lemma 8 *Under Assumption 2:*

$$H_T(\widetilde{P}^{\widetilde{\mu}} || P^\mu) \leq \frac{1}{T} R(\widetilde{\mu} || \mu) + \int \left(\frac{1}{2T}\int_0^T E_{\widetilde{P}}\left[\|\beta(\widetilde{X}_s^x)\|^2\right] ds\right) \widetilde{\mu}(dx).$$

6.3 Example: Euler-Maruyama Methods for SDEs with Additive Noise

As the final example, we consider SDEs with additive noise, approximated by a (generalized) Euler-Maruyama (EM) method.

Assumption 3 *Let W_t be an n-dimensional Wiener process on filtered probability spaces satisfying the usual conditions, $b : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the linear boundedness and local Lipschitz properties as described in Remark 6, and X_t^x be the strong solutions to the SDEs*

$$dX_t^x = b(X_t^x)dt + dW_t, \ X_0^x = x. \tag{117}$$

Recall that versions can be chosen so that X_t^x is jointly continuous in (t, x) and X_t^x satisfies the flow property Eq. (112).

We fix $\Delta t > 0$ and assume we are given a measurable vector field $\widetilde{b}_{\Delta t} : \mathbb{R}^n \to \mathbb{R}^n$ (the drift for the generalized EM method). We define the approximating process $\widetilde{X}_0^x = x$,

$$\widetilde{X}^x|_{(j\Delta t,(j+1)\Delta t]}(t) = \widetilde{X}_{j\Delta t}^x + \widetilde{b}_{\Delta t}(\widetilde{X}_{j\Delta t}^x)(t - j\Delta t) + W_t - W_{j\Delta t} \ for \ j \in \mathbb{Z}_0. \tag{118}$$

We emphasize that, for the purposes of employing the theory we have developed, it is necessary to extend $\widetilde{X}_t^x$ to all $t \geq 0$, and not just define it at the mesh points $j\Delta t$.

Let P denote the probability measure on the space where the SDE is defined. Similarly to the previous example, we define $P^x = (X^x)_* P$ and $\widetilde{P}^x = (\widetilde{X}^x)_* P$, probability measures on

$$(\Omega, \mathcal{F}, \mathcal{F}_t) = (C([0, \infty), \mathbb{R}^n), \mathcal{B}(C([0, \infty), \mathbb{R}^n)), \sigma(\pi_s, s \leq t)). \tag{119}$$

Assumption 3 is sufficient to guarantee that Assumption 1 holds. The chain rule for relative entropy can be used to obtain

$$R(\widetilde{P}_T^{\widetilde{\mu}} || P_T^\mu) \leq R(\widetilde{\mu} || \mu) + \int R(\widetilde{P}_T^x || P_T^x)\widetilde{\mu}(dx).$$

Let $T = N\Delta t$ for $N \in \mathbb{Z}^+$. For the purposes of bounding the relative entropy term

$$R(\widetilde{P}_T^x || P_T^x) = R((\widetilde{X}^x|_{[0,N\Delta t]})_* P || (X^x|_{[0,N\Delta t]})_* P), \tag{120}$$

it will be useful to define the Polish space $\mathcal{Y} \equiv C([0, \Delta t], \mathbb{R}^n)$ and the following one step transition probabilities for a discrete-time Markov process on $\mathcal{Y}$:

$$q(y, B) = P\left(X^{y(\Delta t)}|_{[0,\Delta t]} \in B\right), \ \widetilde{q}(y, B) = P\left(\widetilde{X}^{y(\Delta t)}|_{[0,\Delta t]} \in B\right). \tag{121}$$

Letting $\otimes_1^N q$ denote the composition on $\mathcal{Y}^N$, the Markov property implies

$$\otimes_1^N q(x, \cdot) = \left(X^x|_{[0,\Delta t]}, X^x|_{\Delta t+[0,\Delta t]}, ..., X^x|_{(N-1)\Delta t+[0,\Delta t]}\right)_* P \tag{122}$$

for all $x \in \mathbb{R}^n$, and similarly for $\widetilde{q}$, $\widetilde{X}^x$.

Therefore, the chain rule for relative entropy gives

$$R(\widetilde{P}^x_{N\Delta t}||P^x_{N\Delta t}) = \sum_{j=0}^{N-1} \int R\left(\widetilde{q}(y,\cdot)||q(y,\cdot)\right)\widetilde{q}^j(x,dy). \qquad (123)$$

for all $x \in \mathbb{R}^n$. Hence we arrive at:

Lemma 9

$$R(\widetilde{P}^x_{N\Delta t}||P^x_{N\Delta t}) = \sum_{j=1}^{N} E_P\left[R\left(\widetilde{P}^{(\cdot)}_{\Delta t}||P^{(\cdot)}_{\Delta t}\right) \circ \widetilde{X}^x_{(j-1)\Delta t}\right]. \qquad (124)$$

The one step relative entropy can be bounded via Girsanov's theorem, similarly to Lemma 8; on each time interval of length Δt, the tilde process is simply the solution to an SDE with constant drift and additive noise.

Lemma 10 *Under Assumption 3*

$$H_{N\Delta t}(\widetilde{P}^{\widetilde{\mu}}||P^\mu) \leq \frac{1}{N\Delta t}R(\widetilde{\mu}||\mu) \qquad (125)$$

$$+ \frac{1}{N}\sum_{j=1}^{N}\int\int E_P\left[\frac{1}{2\Delta t}\int_0^{\Delta t}\|\widetilde{b}_{\Delta t}(y) - b(\widetilde{X}^x_s)\|^2 ds\right]\widetilde{p}^{\Delta t}_{j-1}(x,dy)\widetilde{\mu}(dx),$$

where $\widetilde{p}^{\Delta t}_j(x,dy) = (\widetilde{X}^x_{j\Delta t})_* P$.

6.3.1 Euler-Maruyama Error Bounds

We end this section by specializing the results to the Euler-Maruyama method, $\widetilde{b}_{\Delta t} \equiv b$.

If we assume b is C^1 with bounded first derivative and Db is L-Lipschitz then Taylor expanding b gives

$$\int_0^{\Delta t} E_P\left[\|\widetilde{b}_{\Delta t}(y) - b(\widetilde{X}^y_s)\|^2\right]ds \leq \operatorname{tr}\left(Db(y)Db(y)^T\right)\frac{\Delta t^2}{2} \qquad (126)$$

$$+ \|Db(y)b(y)\|^2\frac{\Delta t^3}{3} + \frac{16\sqrt{2}\Gamma((n+3)/2)}{5\Gamma(n/2)}L\|Db\|_\infty \Delta t^{5/2}$$

$$+ \frac{2n(n+2)L^2}{3}\Delta t^3 + L\|Db\|_\infty\|b(y)\|^3\Delta t^4 + \frac{2L^2}{5}\|b(y)\|^4\Delta t^5,$$

and therefore

$$
\begin{aligned}
H_{N\Delta t} \leq & \frac{1}{N\Delta t} R(\widetilde{\mu}||\mu) + \frac{\Delta t}{4} \frac{1}{N} \sum_{j=1}^{N} \int E_P \left[\|Db(\widetilde{X}_{(j-1)\Delta t}^x)\|_F^2 \right] \widetilde{\mu}(dx) \quad (127) \\
& + \Delta t^{3/2} \left(\frac{8\sqrt{2}\Gamma((n+3)/2)}{5\Gamma(n/2)} L\|Db\|_\infty + \frac{n(n+2)L^2}{3} \Delta t^{1/2} \right. \\
& + \frac{1}{N} \sum_{j=1}^{N} \int E_P \left[\frac{\Delta t^{1/2}}{6} \|Db(\widetilde{X}_{(j-1)\Delta t}^x)b(\widetilde{X}_{(j-1)\Delta t}^x)\|^2 \right. \\
& + \left. \left. \frac{L\|Db\|_\infty}{2} \|b(\widetilde{X}_{(j-1)\Delta t}^x)\|^3 \Delta t^{3/2} + \frac{L^2}{5} \|b(\widetilde{X}_{(j-1)\Delta t}^x)\|^4 \Delta t^{5/2} \right] \widetilde{\mu}(dx) \right),
\end{aligned}
$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm.

This isn't the tightest possible bound and alternatives can be obtained by Taylor expanding further, but they give an idea of the type of result that can be obtain under various smoothness assumptions on b.

If $d\widetilde{\mu} = e^{-\widetilde{\phi}}dx$ and $d\mu = e^{-\phi}dx$ where $\widetilde{\phi}$ and ϕ are known functions then the relative entropy term takes the form:

$$
R(\widetilde{\mu}||\mu) = \int (\phi(x) - \widetilde{\phi}(x))e^{-\widetilde{\phi}(x)}dx. \quad (128)
$$

Assuming that one can sample from $\widetilde{\mu}$, Eq. (128) and Eq. (127) can estimated via Monte Carlo methods, providing UQ bounds that involve a mixture of a priori and a posteriori data.

A Proof of the Perturbation Bound

Lemma A1 *Let H be a Hilbert space, $A : D(A) \subset H \to H$ a linear operator and $B : H \to H$ a bounded self-adjoint operator. Suppose there exists $D > 0$ and $x_0 \in H$ with $\|x_0\| = 1$ such that*

$$
\langle Bx_0, x_0 \rangle = 0 \quad and \quad \mathrm{Re}(\langle Ax, x \rangle) \leq -D\|P^\perp x\|^2 \quad (129)
$$

for all $x \in D(A)$, where $P^\perp$ is the orthogonal projector onto $x_0^\perp$.

Define

$$
B^+ \equiv \max \left\{ \sup_{\|y\|=1} \langle By, y \rangle, 0 \right\}. \quad (130)
$$

Then for any $0 \leq c < D/B^+$ we have

$$
\sup_{x \in D(A), \|x\|=1} \mathrm{Re}(\langle (A + cB)x, x \rangle) \leq \frac{c^2\|Bx_0\|^2}{D - cB^+}. \quad (131)
$$

Proof Let $x \in D(A)$ with $\|x\| = 1$. Define $a = \langle x_0, x \rangle$. (Here we will use the convention of linearity in the second argument). We have $\|P^\perp x\|^2 = 1 - |a|^2$, and so $|a| \leq 1$ with equality if and only if $P^\perp x = 0$.

We can decompose $x = ax_0 + \sqrt{1 - |a|^2}v$, where either $v = 0$ and $|a| = 1$ if $P^\perp x = 0$ or $v = P^\perp x/\sqrt{1 - |a|^2}$ and $\|v\| = 1$ if $P^\perp x \neq 0$. Either way, $v \perp x_0$.

With this, we have

$$\text{Re}(\langle(A + cB)x, x\rangle) = \text{Re}(\langle Ax, x\rangle) + c\,\text{Re}(\langle Bx, x\rangle) \tag{132}$$

$$\leq - D(1 - |a|^2) + 2c\,\text{Re}(\langle\sqrt{1 - |a|^2}v, aBx_0\rangle) + c(1 - |a|^2)\langle Bv, v\rangle$$

$$\leq 2c|a|\sqrt{1 - |a|^2}\|Bx_0\| - \left(D - cB^+\right)(1 - |a|^2),$$

where B^+ is given by Eq. (130).

Restricting to $0 \leq c < D/B^+$, we can compute

$$\sup_{x \in D(A), \|x\|=1} \text{Re}(\langle(A + cB)x, x\rangle) \leq \sup_{r \geq 0}\left(2c\|Bx_0\|r - \left(D - cB^+\right)r^2\right) = \frac{\|Bx_0\|^2 c^2}{D - cB^+}. \tag{133}$$

$\square$

The previous lemma is closest in spirit to the probabilistic application, as $\|Bx_0\|^2$ plays the role of the variance. However, one can work with non-self-adjoint perturbations, if one instead uses the definition

$$B^+ \equiv \max\left\{\sup_{\|y\|=1} \text{Re}(\langle By, y\rangle), 0\right\} \tag{134}$$

and makes the replacement $\|Bx_0\| \to (\|Bx_0\| + \|B\|)/2$ in Eq. (131). The proof is similar.

B F-Sobolev Inequalities

Theorem 11 can be generalized to the F-Sobolev case; see [13] for a proof of the following:

Theorem B1 *Let A be the generator of $\mathcal{P}_t$ and μ^* be an invariant measure. Suppose we have a function $F : (0, \infty) \to \mathbb{R}$ satisfying:*

1. *F is strictly increasing,*
2. *F is concave (hence continuous),*
3. *$F(1) = 0$,*
4. *$F(x) \to \infty$ as $x \to \infty$,*
5. *$F(xy) \leq F(x) + F(y)$ for all $x, y \geq 0$.*

(Note that this implies $F^{-1} : (F(0^+), \infty) \to (0, \infty)$ exists, is increasing, convex, and continuous.)

Assume the F-Sobolev inequality holds for μ^*:

$$\int g^2 F(g^2)d\mu^* \leq -\int A[g]g\,d\mu^* \quad \text{for all } g \in D(A, \mathbb{R}) \text{ with } \|g\|_{L^2(\mu^*)} = 1. \tag{135}$$

Finally, suppose that $V \in L^1(\mu^*)$ with $V > F(0^+)$ and $\int F^{-1}(V)d\mu^* < \infty$. Then $\mathcal{P}_t^V : L^2(\mu^*) \to L^2(\mu^*)$, defined by

$$\mathcal{P}_t^V[g](x) = E^x\left[g(X_t)\exp\left(\int_0^t V(X_s)ds\right)\right], \tag{136}$$

are well-defined linear operators and the operator norm satisfies the bound

$$\|\mathcal{P}_t^V\| \leq \exp\left[tF\left(\int F^{-1}(V)d\mu^*\right)\right]. \tag{137}$$

Note that if $F(0^+) = -\infty$ then certain unbounded observables are allowed, namely those that satisfy the integrability condition Eq. (92).

This theorem leads to a UQ bound of the form, Eq. (34). The proof is analogous to the log-Sobolev case from Section 4.4.

Theorem B2 *In addition to Assumption 1, assume the F-Sobolev inequality, Eq. (135), holds for some function, F, having the properties listed in Theorem B1, $f \in L^1(\mu^*, \mathbb{R})$, and there exists $c_- < 0 < c_+$ such that, for all $c \in (c_-, c_+)$:*

$$F(0^+) < \pm c(f - \mu^*[f]), \quad \int F^{-1}\left(\pm c\left(f - \mu^*[f]\right)\right) d\mu^* < \infty. \tag{138}$$

Then a UQ bound of the form Eq. (34) holds with

$$\Lambda(c) = \begin{cases} F\left(\int F^{-1}(V_c)d\mu^*\right) & \text{if } c \in (c_-, c_+) \\ +\infty & \text{otherwise} \end{cases}. \tag{139}$$

In addition, if $\mathrm{Var}_{\mu^*}[f] > 0$, F and F^{-1} are smooth, $F'(1) > 0$, $(F^{-1})''(0) > 0$, and $c \to \mu^*[F^{-1}(V_c)]$ is smooth on a neighborhood of 0 and can be differentiated under the integral then Eq. (26) holds with

$$\Lambda''(0) = F'(1)(F^{-1})''(0)\,\mathrm{Var}_{\mu^*}[f], \quad \eta = \frac{1}{T}R(\widetilde{P}_T^{\widetilde{\mu}}\|P_T^{\mu^*}). \tag{140}$$

C Continuous-Time Jump Processes on General State Spaces

In this appendix we generalize the construction of continuous-time, pure-jump processes from countable state spaces to general state spaces. This will allow us to construct a continuous-time Markov process whose semigroup is $\mathcal{P}_t = \exp(t(\mathcal{P} - I))$, where $\mathcal{P}$ is a discrete-time semigroup operator, as well as compute the relative entropy of two such processes; this is needed in Section 5. The construction closely mirrors the discrete state-space case as found in, for example, Appendix 1 in [33].

Let $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ be a Polish space and $p(x, dy)$ be a probability kernel on $\mathcal{X}$. Given $\lambda > 0$ define the probability kernel, p^J, on the Polish space $(\mathcal{X} \times (0, \infty), \mathcal{B}_{\mathcal{X}} \bigotimes \mathcal{B}_{(0,\infty)})$:

$$p^J((x, s), \cdot) = p(x, dy) \times \lambda e^{-\lambda t} dt. \tag{141}$$

For any probability measure π on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$, let P^π (for $\pi = \delta_x$ we simply write P^x) be the unique probability measure on $(\prod_{n=0}^{\infty}(\mathcal{X} \times (0, \infty)), \bigotimes_{n=0}^{\infty}(\mathcal{B}_{\mathcal{X}} \bigotimes \mathcal{B}_{(0,\infty)}))$ generated by the transition probabilities p^J and initial distribution $\pi \times (\lambda e^{-\lambda t} dt)$. Define

$$\Omega = \left\{ \{(x_i, s_i)\}_{i=0}^{\infty} \in \prod_{n=0}^{\infty}(\mathcal{X} \times (0, \infty)) : \sum_{i=0}^{\infty} s_i = \infty \right\}. \tag{142}$$

Then $P^\pi(\Omega) = 1$ and, working on the probability space $(\Omega, \mathcal{F}, P^\pi|_\Omega)$ (from here on, we simply write P^π for $P^\pi|_\Omega$), where $\mathcal{F} \equiv \left(\bigotimes_{n=0}^{\infty}(\mathcal{B}_{\mathcal{X}} \bigotimes \mathcal{B}_{(0,\infty)})\right) \cap \Omega$, define the jump process, jump intervals, and jump times on Ω:

$$X_n^J \equiv \pi_1 \circ \pi_n, \quad \Delta_n \equiv \mu \circ \pi_n, \text{ for } n \in \mathbb{Z}_0, \quad J_0 \equiv 0, \quad J_n \equiv \sum_{k=0}^{n-1} \Delta_k \text{ for } n \in \mathbb{Z}^+, \tag{143}$$

where π_i denote projections onto components. Note that $J_n(\omega) \to \infty$ as $n \to \infty$ for all $\omega \in \Omega$.

(X_n^J, Δ_n) is a Markov process on (Ω, P^π) with transition probabilities p^J and initial distribution $\pi \times (\lambda e^{\lambda t} dt)$.

Define the càdlàg process

$$X_t(\omega) = X_n^J, \text{ where } t \in [J_n(\omega), J_{n+1}(\omega)) \tag{144}$$

($J_n(\omega) \to \infty$ for all ω, so Ω is a disjoint union of $\{J_n \leq t < J_{n+1}\}$, $n \geq 0$) and let $\mathcal{F}_t$ be the natural filtration for X_t.

Finally, define the probability kernels on $\mathcal{X}$

$$p_t(x, A) \equiv P^x(X_t \in A), \ t \geq 0, \ x \in \mathcal{X}. \tag{145}$$

With this setup, we have the following:

Theorem C3 $(\Omega, \mathcal{F}, \mathcal{F}_t, X_t, P^x)$, $x \in \mathcal{X}$, is a càdlàg Markov family with transition probabilities p_t (see the definition in Assumption 1).

One also obtains realizability of the semigroup $\exp(t\lambda(\mathcal{P} - I))$ by a probability kernel:

Theorem C4 If μ^* is invariant measure for p then μ^* is invariant for p_t for all $t \geq 0$ and the bounded linear operators on $L^2(\mu^*)$,

$$\mathcal{P}[f](x) \equiv \int f(y)p(x, dy), \ \ \mathcal{P}_t[f](x) \equiv \int f(y)p_t(x, dy), \tag{146}$$

satisfy

$$\mathcal{P}_t = \exp(t\lambda(\mathcal{P} - I)), \tag{147}$$

for all $t \geq 0$, where right side is the operator exponential for bounded operators on $L^2(\mu^*)$.

These results are all straightforward to prove by using the same strategy as the discrete-time case; specifically, for any $t \geq 0$, decompose Ω into a disjoint union of $\{J_n \leq t < J_{n+1}\}$, and then expand quantities in terms of the transition probabilities p^J and use the following lemma:

Lemma C2

$$P^\pi(J_{n+1} > t | X_k^J, J_k : k \leq n) = e^{-\lambda(t - J_n)}1_{J_n \leq t} + 1_{J_n > t}. \tag{148}$$

The proof of this lemma closely mirrors the corresponding proof in the discrete state-space case.

We now derive a formula for the Radon-Nikodym derivative for two measures constructed as above. Note that the jump chain (X_n^J, Δ_n) is generally not recoverable from X_t; specifically, the J_n are not $\mathcal{F}_t$-stopping times (this is because 'jumps' don't necessarily change the state, unlike the construction commonly used when the state space is discrete). Hence, we derive a formula for the Radon-Nikodym derivative on the enlarged filtration

$$\mathcal{G}_t \equiv \sigma(1_{J_n \leq s}, X_{J_n \wedge s} : s \leq t, n \geq 0). \tag{149}$$

We will also need the following lemmas, which are simple to prove by using similar strategies to the one described above:

Lemma C3 Let $F : (0, \infty)^n \times \mathcal{X}^{n+1} \to \mathbb{R}$ be measurable and non-negative (i.e. $F \in L^+$). Then for any $t > 0$, $n \geq 1$ we have

$$E^\pi[F(\Delta_0, ..., \Delta_{n-1}, X_0^J, ..., X_n^J)] = E^\pi[\widetilde{F}(\Delta_0, ..., \Delta_{n-1}, X_0^J, ..., X_{n-1}^J)] \tag{150}$$

where

$$\widetilde{F}(s_0, ..., s_{n-1}, x_0, ..., x_{n-1}) \equiv \int F(s_0, ..., s_{n-1}, x_0, ..., x_{n-1}, z)p(x_{n-1}, dz). \tag{151}$$

Lemma C4 Let $F \in L^+(\mathcal{X})$. Then for any $t \geq 0$ we have

$$\sum_{n \geq 0} E^\pi\left[F(X_n^J)1_{J_{n+1} \leq t}\right] = \lambda E^\pi\left[\int_0^t F(X_s)ds\right]. \tag{152}$$

Lemma C5 *Suppose we have probability measures $\widetilde{\mu}$, μ and probability kernels $\widetilde{p}(x, dy)$, $p(x, dy)$ on $\mathcal{X}$. Assume that $\widetilde{\mu} \ll \mu$ and $\widetilde{p}(x, \cdot) \ll p(x, \cdot)$ for $\widetilde{\mu}$ a.e. x. Recall that this implies existence of $h \in L^+(\mathcal{X} \times \mathcal{X})$ such that*

$$\widetilde{p}(x, dy) = h(x, y)p(x, dy) \text{ for } \widetilde{\mu} \text{ a.e. } x. \tag{153}$$

Also suppose that $(\widetilde{\mathcal{P}}^*)^n[\widetilde{\mu}] \ll \widetilde{\mu}$ for all n. Then $\widetilde{\mu} \otimes_1^n \widetilde{p} \ll \mu \otimes_1^n p$ for all n and

$$\widetilde{\mu} \otimes_1^n \widetilde{p}(dx_0, ..., dx_n) = \frac{d\widetilde{\mu}}{d\mu}(x_0) \prod_{i=1}^n h(x_{i-1}, x_i)\mu \otimes_1^n p(dx_0, ..., dx_n). \tag{154}$$

Now we compute the Radon-Nikodym derivative:

Theorem C5 *Suppose we have probability measures $\widetilde{\mu}$, μ and probability kernels $\widetilde{p}(x, dy)$, $p(x, dy)$ on $\mathcal{X}$. Assume that $\widetilde{\mu} \ll \mu$ and $\widetilde{p}(x, \cdot) \ll p(x, \cdot)$ for $\widetilde{\mu}$ a.e. x. Given $\lambda > 0$, construct the probability measures $\widetilde{P}^{\widetilde{\mu}}$ and P^μ on Ω from $\widetilde{p}$ and p respectively, and define the process X_t as in Eq. (144).*

Suppose $(\widetilde{\mathcal{P}}^*)^n[\widetilde{\mu}] \ll \widetilde{\mu}$ for all n (in particular, if $\widetilde{\mu}$ is invariant for $\widetilde{p}$). Then for any $t \geq 0$ we have $\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t} \ll P^\mu|_{\mathcal{G}_t}$ and

$$\frac{d\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t}}{dP^\mu|_{\mathcal{G}_t}} = \frac{d\widetilde{\mu}}{d\mu}(X_0) \prod_{n \geq 1: J_n \leq t} h(X_{J_{n-1} \wedge t}, X_{J_n \wedge t}), \tag{155}$$

where h was defined in Eq. (153).

Proof

$$G \equiv \frac{d\widetilde{\mu}}{d\mu}(X_0) \prod_{n \geq 1: J_n \leq t} h(X_{J_{n-1} \wedge t}, X_{J_n \wedge t}) \tag{156}$$

is $\mathcal{G}_t$-measurable, so it suffices to show that

$$\widetilde{E}^{\widetilde{\mu}} \left[F(1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j} : j = 1, ..., m) \right] = E^\mu \left[F(1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j} : j = 1, ..., m)G \right] \tag{157}$$

for all choices of m, $s_j \leq t$, $n_j \geq 0$, $F \in L^+$.

We have

$$\widetilde{E}^{\widetilde{\mu}} \left[F(1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j} : j = 1, ..., m) \right] \tag{158}$$

$$= \sum_{n \geq 0} \widetilde{E}^{\widetilde{\mu}} \left[1_{J_n \leq t < J_{n+1}} F(1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j} : j = 1, ..., m) \right].$$

On $J_n \leq t < J_{n+1}$ one can write

$$(1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j} : j = 1, ..., m) = \psi(X_\ell^J, J_\ell : \ell \leq n), \tag{159}$$

where ψ is measurable (ψ depends on n and the n_j, s_j).

Therefore

$$\widetilde{E}^{\widetilde{\mu}} \left[F(1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j} : j = 1, ..., m) \right] \tag{160}$$

$$= \sum_{n \geq 0} \widetilde{E}^{\widetilde{\mu}} \left[1_{J_n \leq t < J_{n+1}} F(\psi(X_\ell^J, J_\ell : \ell \leq n)) \right]$$

$$= \sum_{n \geq 0} \widetilde{E}^{\widetilde{\mu}} \left[1_{J_n \leq t} e^{-\lambda(t - J_n)} F(\psi(X_\ell^J, J_\ell : \ell \leq n)) \right].$$

Similarly

$$E^\mu \left[F(\prod_{j=1}^m (1_{J_{n_j} \leq s_j}, X_{J_{n_j} \wedge s_j})) G \right] \qquad (161)$$

$$= \sum_{n \geq 0} E^\mu \left[1_{J_n \leq t} e^{-\lambda(t-J_n)} F(\psi(X_\ell^J, J_\ell : \ell \leq n)) \frac{d\widetilde{\mu}}{d\mu}(X_0) \prod_{k=1}^n h(X_{k-1}^J, X_k^J) \right].$$

We can then use Lemma C5 to get

$$\widetilde{E}^{\widetilde{\mu}} \left[1_{J_n \leq t} e^{-\lambda(t-J_n)} F(\psi(X_\ell^J, J_\ell : \ell \leq n)) \right] \qquad (162)$$

$$= \int \int 1_{\sum_{\ell=0}^{n-1} s_\ell \leq t} e^{-\lambda(t-\sum_{\ell=0}^{n-1} s_\ell)} F(\psi(x_k, \sum_{\ell=0}^{k-1} s_\ell : k \leq n)) \frac{d\widetilde{\mu}}{d\mu}(x_0) \prod_{i=1}^n h(x_{i-1}, x_i)$$

$$\times \mu \otimes_1^n p(dx_0, ..., dx_n) \prod_{i=0}^n (\lambda e^{-\lambda s_i} ds_i)$$

$$= E^\mu \left[1_{J_n \leq t} e^{-\lambda(t-J_n)} F(\psi(X_k^J, J_k : k \leq n)) \frac{d\widetilde{\mu}}{d\mu}(X_0) \prod_{i=1}^n h(X_{i-1}^J, X_i^J) \right],$$

which proves the claim. □

Combining Theorem C5 with Lemmas C3 and C4 we obtain a formula for the relative entropy.

Theorem C6 *Suppose we have probability measures $\widetilde{\mu}$, μ and probability kernels $\widetilde{p}(x, dy)$, $p(x, dy)$ on $\mathcal{X}$. Assume that $\widetilde{\mu} \ll \mu$ and $\widetilde{p}(x, \cdot) \ll p(x, \cdot)$ for $\widetilde{\mu}$ a.e. x.*

Suppose $(\widetilde{\mathcal{P}}^*)^n[\widetilde{\mu}] \ll \widetilde{\mu}$ for all n (in particular, if $\widetilde{\mu}$ is invariant for $\widetilde{p}$). Then for any $t \geq 0$

$$R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t} || P^\mu|_{\mathcal{G}_t}) = R(\widetilde{\mu}||\mu) + \lambda \int_0^t \widetilde{E}^{\widetilde{\mu}} \left[\int \log(h(X_s, z)) h(X_s, z) p(X_s, dz) \right] ds, \qquad (163)$$

where h was defined in Eq. (153).

It is also useful to note that $\mathcal{F}_t \subset \mathcal{G}_t$ implies

$$R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{F}_t} || P^\mu|_{\mathcal{F}_t}) \leq R(\widetilde{P}^{\widetilde{\mu}}|_{\mathcal{G}_t} || P^\mu|_{\mathcal{G}_t}). \qquad (164)$$

Corollary C1 *Suppose we have probability measures $\widetilde{\mu}^*$, μ and probability kernels $\widetilde{p}(x, dy)$, $p(x, dy)$ on $\mathcal{X}$. If $\widetilde{\mu}^*$ is invariant for $\widetilde{p}$ then for all $t > 0$*

$$R(\widetilde{P}^{\widetilde{\mu}^*}|_{\mathcal{F}_t} || P^\mu|_{\mathcal{F}_t}) \leq R(\widetilde{\mu}^*||\mu) + \lambda t \int R(\widetilde{p}(x, \cdot) || p(x, \cdot)) d\widetilde{\mu}^*. \qquad (165)$$

Acknowledgments

References

1. K. Chowdhary and P. Dupuis. Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(3):635662, 2013.
2. P. Dupuis, M.A. Katsoulakis, Y. Pantazis, and P. Plecháč. Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):80–111, 2016.
3. Y. Pantazis and M.A. Katsoulakis. A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics. *The Journal of Chemical Physics*, 138(5):054115, 2013.
4. K. Gourgoulias, M.A. Katsoulakis, and L. Rey-Bellet. Information Metrics For Long-Time Errors in Splitting Schemes For Stochastic Dynamics and Parallel Kinetic Monte Carlo. *SIAM Journal on Scientific Computing*, 38(6):A3808–A3832, 2016.
5. M.A. Katsoulakis, L. Rey-Bellet, and J. Wang. Scalable information inequalities for uncertainty quantification. *Journal of Computational Physics*, 336:513 – 545, 2017.
6. P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and L. Rey-Bellet. Sensitivity analysis for rare events based on Rényi divergence. *ArXiv e-prints*, May 2018.
7. K. Gourgoulias, M.A. Katsoulakis, L. Rey-Bellet, and J. Wang. How biased is your model? concentration inequalities, information and model bias. *ArXiv e-prints*, 2017.
8. T. Breuer and I. Csiszár. Systematic stress tests with entropic plausibility constraints. *Journal of Banking & Finance*, 37(5):1552 – 1559, 2013.
9. H. Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016.
10. P. Glasserman and X. Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014.
11. L. Wu. A deviation inequality for non-reversible Markov processes. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 36(4):435 – 445, 2000.
12. P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov Processes. *ESAIM: Probability and Statistics*, 5:pp 183–201, December 2001. http://www.sherpa.ac.uk/romeo/issn/1292-8100/.
13. P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM: Probability and Statistics*, 12:1229, 2008.
14. A. Guillin, C. Léonard, L. Wu, and N. Yao. Transportation-information inequalities for Markov processes. *Probability Theory and Related Fields*, 144(3):669–695, Jul 2009.
15. F. Gao, A. Guillin, and L. Wu. Bernstein-type Concentration Inequalities for Symmetric Markov Processes. *Theory of Probability & Its Applications*, 58(3):358–382, 2014.
16. A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2009.
17. P. Dupuis and R.S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics. Wiley, 2011.
18. S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013.
19. T. Breuer and I. Csiszár. Measuring distribution model risk. *Mathematical Finance*, 26(2):395–411.
20. K. Yosida. *Functional Analysis, 3rd edn.* Springer, 1971.
21. D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures. *Electron. Commun. Probab.*, 13:60–66, 2008.
22. D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *Journal of Functional Analysis*, 254(3):727 – 759, 2008.
23. P. Cattiaux. Long time behavior of Markov processes. *ESAIM: Proc.*, 44:110–128, 2014.
24. M. Hairer. Convergence of Markov processes, 2010.
25. M. Hairer and J. Mattingly. Yet another look at Harris ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117. 2011.

26. D. Chafaï. Binomial-Poisson entropic inequalities and the M/M/∞ queue. *ESAIM: PS*, 10:317–339, 2006.
27. O. S. Rothaus. Lower bounds for eigenvalues of regular Sturm-Liouville operators and the logarithmic Sobolev inequality. *Duke Math. J.*, 45(2):351–362, 06 1978.
28. E.A. Carlen and M. Loss. Logarithmic Sobolev inequalities and spectral gaps. In *Recent Advances in the Theory and Applications of Mass Transport. Contemp. Math., vol. 353*, page 5360. 2004.
29. D. Bakry and M. Emery. Hypercontractivité do semi-groups de diffusion. *C.R. Acad. Sci. Paris Sér I Math*, 299:775–778, 1984.
30. D.A. Levin and Y. Peres. *Markov Chains and Mixing Times: Second Edition*. MBK. American Mathematical Society, 2017.
31. P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 08 1996.
32. P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *The Annals of Applied Probability*, 3(3):696–730, 1993.
33. C. Kipnis and C. Landim. *Scaling Limits of Interacting Particle Systems*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2013.
34. I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 2014.