Journal of Computational Physics 330 (2017) 127–133

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

An estimator for the relative entropy rate of path measures for stochastic differential equations

Manfred Opper

Artificial Intelligence Group, Technische Universität Berlin, Marchstraße 23, Berlin 10587, Germany

ARTICLE INFO

Article history:
Received 19 May 2016
Received in revised form 1 November 2016
Accepted 13 November 2016
Available online 17 November 2016

Keywords:
Stochastic differential equations
Relative entropy
Drift estimation
Variational bound

ABSTRACT

We address the problem of estimating the relative entropy rate (RER) for two stochastic processes described by stochastic differential equations. For the case where the drift of one process is known analytically, but one has only observations from the second process, we use a variational bound on the RER to construct an estimator.

1. Introduction

The Kullback–Leibler (KL) divergence or relative entropy plays an important role as an error measure in approximating one probability distribution by another. It appears implicitly in the free energy bound of the *variational approximation* in statistical mechanics. The variational approach has been adopted within the field of machine learning where it is used to approximate intractable posterior distributions in Bayesian inference. More recently, KL-divergence related cost functions have found applications to infinite dimensional path measures of stochastic processes. This includes e.g. work on optimal control problems [1–3], probabilistic inference [4,5] and rare event simulations [6]. The relative entropy rate (RER), i.e. the large time limit of the KL-divergence per time, plays an important role for uncertainty quantification and sensitivity analysis [7–9] of stochastic dynamics with applications to (bio)chemical reaction networks (see e.g. [10]). Finally, the RER has been used [11,12] to quantify the loss of information when a non-equilibrium model in statistical physics is approximated by a coarse-grained dynamics.

In this paper we address the problem of estimating the RER for stochastic differential equations from observations of the process. This rate can be expressed by the drift functions of the two processes and the stationary density of one of them. Its evaluation is nontrivial, when one path distribution (usually the one used for approximation) is characterised by an analytical expression for the drift, but the other distribution only via observations of the process.

The main contribution of the paper is a simple estimator (eq. (22)) for the RER which is valid for processes which satisfy a specific potential condition for the difference between the two drifts. The estimator is derived from a variational representation of the RER (eq. (11)), which involves a functional that is an expectation with respect to the stationary distribution of the process and which can be estimated from ergodic samples.

E-mail address: manfred.opper@tu-berlin.de.

http://dx.doi.org/10.1016/j.jcp.2016.11.021
0021-9991/

Our work was motivated by estimators for log-probability densities [13,14] developed in the field of machine learning. Using a related variational method, we have constructed estimators for drift functions from the empirical distribution of data [15]. The application of these ideas to the estimation of relative entropy rates given in this paper is new.

The paper is organised as follows: Section two introduces relative entropy rates for stochastic differential equations. Section 3 gives a variational bound on the RER which depends on the stationary density of the process. Replacing this density by the empirical distribution of observations in section 4 is used to get a simple estimator within a parametric framework. Section 5 discusses the relation of the potential condition to optimal stochastic control. The performance of the estimator is illustrated on a simple one-dimensional double well problem in section 6. We conclude with an outlook and discussion in section 7. In the appendix we give a proof of a stochastic representation of our cost function.

2. The relative entropy rate for stochastic differential equations

We consider stochastic differential equations (SDE) for the dynamics of a d-dimensional diffusion process $X_t \in R^d$ given by

$$dX_t = g(X_t)dt + \sigma(X_t)dW_t\,. \tag{1}$$

The drift function $g(\cdot) \in R^d$ represents the deterministic part of the driving force and W is a k-dimensional ($k \le d$) vector of independent Wiener processes acting as a white noise source. The strength of the noise is determined by the state dependent $d \times k$ dimensional diffusion matrix $\sigma(\cdot)$.

The relative entropy or Kullback–Leibler divergence between the probability measures $P^g(X_{0:T})$ and $P^f(X_{0:T})$ for paths $X_{0:T}$ of two processes with drifts $g(x)$ and $f(x)$ is given by (see e.g. [4,12])

$$D_T(P^g, P^f) \doteq E_{P^g}\left[\ln \frac{dP^g}{dP^f}\right] = \frac{1}{2}\int_0^T dt \int p_t^g(x)\|g(x) - f(x))\|^2_{D^{-1}}\,.$$

Here $D(x) \doteq \sigma(x)\sigma(x)^\top$ is the diffusion matrix for both processes. $p_t^g(x)$ is the marginal density of the process P^g at time t and the square norm is defined as $\|u(x)\|_A^2 \doteq u(x)\cdot A(x)u(x)$ for positive definite matrices A. Assuming that $p_t^g(x)$ converges to the stationary density $p^g(x)$ for $t \to \infty$, we consider the *relative entropy rate* (RER)

$$d(P^g, P^f) \doteq \lim_{T\to\infty} \frac{1}{T} D_T(P^g, P^f) = \frac{1}{2}\int p^g(x)\|g(x) - f(x))\|^2_{D^{-1}} dx\,. \tag{2}$$

Assume that the drift $f(x)$ and the diffusion $D(x)$ are known, but we don't know g. We would like to estimate the RER using observations generated by the process P^g over a large time window T.

For general drift functions g, following (2), one would have to estimate both the drift and the stationary density $p^g(x)$ from the observations. This double estimation problem simplifies when $D = \sigma^2 I$ and the drift is derived from a potential i.e. $g(x) = \sigma^2\nabla\psi(x)$. Then the stationary density is given by $p^g(x) \propto e^{2\psi(x)}$ and one might use a smooth estimator of the density $p^g(x)$, e.g. a kernel density estimator (KDE) in order to estimate the RER. One drawback of this method is the fact that the accuracy of the KDE is expected to deteriorate rapidly with increasing dimensionality (see e.g. [14]).

We will next introduce a different approach which avoids the explicit estimate of the density but estimates the difference between g and f. It is based on a variational formulation for the RER (2) and requires a generalised potential condition. To be specific, we assume that the drift is of the form

$$g(x) = f(x) + D(x)\nabla\psi^*(x)\,. \tag{3}$$

This includes the equilibrium case mentioned before but also more general models such as Langevin dynamics [15]. We will later argue in Section 5 that this potential condition naturally appears in the formulation of specific stochastic control problems.

The stationary density $p^g(x)$ for drift (3) is a solution of the Fokker–Planck equation

$$\mathcal{L}_f p^g(x) - \nabla\cdot(D(x)\nabla\psi^*(x)p^g(x)) = 0 \tag{4}$$

where the Fokker–Planck operator $\mathcal{L}_f$ for the drift $f(x)$ is given by

$$\mathcal{L}_f p^g(x) = \nabla\cdot\left[-f(x)p^g(x) + \frac{1}{2}\nabla\cdot(D(x)p^g(x))\right]. \tag{5}$$

The RER (2) becomes

$$d(P^g, P^f) = \frac{1}{2}\int p^g(x)\left\|\nabla\psi^*(x)\right\|^2_D dx\,. \tag{6}$$