Extracting Relevant Physical Parameters from Stochastic Trajectories with Variational Autoencoders

Gabriel Fernández-Fernández* and Alexandre Dauphin

ICFO – Institut de Ciències Fotòniques,

The Barcelona Institute of Science and Technology,

Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain

Carlo Manzo

Facultat de Ciències, Tecnologia i Enginyeries,

Universitat de Vic – Universitat Central de Catalunya (UVic-UCC),

C. de la Laura, 13, 08500 Vic, Spain

Maciej Lewenstein

ICFO – Institut de Ciències Fotòniques,

The Barcelona Institute of Science and Technology,

Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain and

ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

Gorka Muñoz-Gil

Institute for Theoretical Physics, University of Innsbruck,

Technikerstr. 21a, A-6020 Innsbruck, Austria

(Dated: June 29, 2023)

Machine learning (ML) has become an increasingly important tool in scientific research, allowing researchers to analyze large datasets and make accurate forecasts about complex systems. However, understanding why and how the machine makes these predictions is crucial for ensuring their reliability, transparency, and trustworthiness. The field of interpretable ML has developed different approaches to tackle this challenge, which can benefit from the models provided by physics. The exact physical models provide a controlled testing ground for these methods. Additionally, interpreting their predictions can provide valuable insights into the underlying patterns and relationships in the data, and help in the development of new hypotheses or research directions.

Furthermore, statistical physics and ML also have a mutually beneficial relationship. In fact, on the one hand statistical physics has been instrumental in some of the most significant discoveries in ML and, on the other hand, the application of ML methods to the study of stochastic processes has been highly successful. In experimental situations where data is scarce and noisy, ML methods have demonstrated superior performance compared to traditional approaches. An illustrative example is provided by the study of molecular motion using data from single-particle tracking.

In this work, we aim to use interpretable ML methods to extract, in an unsupervised form, the relevant physical parameters of given sets of stochastic trajectories. For this task, we extend the β-variational autoencoder (β-VAE) architecture to properly account for the typical properties found in diffusion data. To show the validity of this approach, I will describe the application of β-VAE to simulated data from stochastic processes such as Brownian motion, fractional Brownian motion, and scaled Brownian motion. Our analysis of the β-VAE reveals its ability to help identify relevant parameters of the data, and generate new trajectories with the learned statistics. Our future perspectives include the application of this technique to other stochastic processes such as continuous time random walks and Ornstein-Uhlenbeck processes, as well as experimental applications where one can extract the minimal representation of trajectories and establish connections with existing theories.

* gabriel.fernandez@icfo.eu