Bridging the gap between stochastic methods and deep learning
Fully funded PhD Position

Institutions: Université Bretagne Sud, LMBA & IRISA Laboratories
Location: Campus Tohannic, Vannes, France
Advisors: François SEPTIER, Nicolas COURTY & Lucas DRUMETZ
Contact: {francois.septier,nicolas.courty}@univ-ubs.fr, lucas.drumetz@imt-atlantique.fr

Subject

Many complex real-world phenomena can be described through probabilistic models that characterize available data, and possibly relate them to other unknown quantities of interest. Fields where such systems can be found include environmental science, biology, econometrics, astronomy, among many others. Unfortunately, for most probabilistic models of practical interest, exact inference is intractable, and so we have to resort to some form of approximation.

Monte Carlo methods are stochastic methods that allow to approximate distributions with a set of random samples [1]. Then, any moments or confidence region with respect to this distribution could be empirically approximated by a discrete sum over these generated samples. Unfortunately, in most cases, sampling directly from the desired distribution is not possible due to its complex nature (multi-modality, high-dimensionality,etc.) or since this distribution is only known up to a normalizing constant (e.g. in Bayesian inference). This has led to the development in recent years of much more advanced algorithms which allow one to obtain the required samples from this target distribution by using either (a) the Markov Chain Monte Carlo (MCMC) methods, which generate a Markov chain whose stationary distribution is the target distribution, or (b) importance sampling (IS) algorithms, where samples are generated from simple proposal densities and are then properly weighted. Despite the existence of theoretical guarantees, convergence speed of such techniques strongly depends on choosing an appropriate proposal distribution, which in practice is quite challenging in practice especially in high-dimensional spaces [2].

On the other hand, deep neural networks have achieved great successes for approximating a deterministic mapping on high-dimensional spaces in a range of different challenging applications related to computer vision, natural language processing, etc. Unfortunately, modern deep learning models used in practice do not capture model uncertainty as they only provide point estimate of parameters and predictions.

The aim of this thesis is to develop theory and methods by successfully combining ideas from both computational statistics and machine learning, thus providing novel stochastic methods to simulate from complex and high dimensional distribution.

We propose to firstly study the current state-of-the-art and more specifically methods that have been recently proposed in the literature, such as *Distilling importance sampling* [3] or *MetFlow* [4] for example. These methods use the principle of normalizing flows, a family of generative models proposed in the ML community [5], in order to efficiently design the proposal distribution of classical sampling techniques. The theory of Normalizing Flows buries a lot of similarities with Optimal Transport [6, 7], for which the supervising team has already a strong expertise [8–10]. The candidate will explore links between the two, and should propose novel methods at the interface of those domains, such as [11]. Then, we will propose novel strategies to be able to deal with high-dimensional spaces as well as flow across different dimensions. An important aspect

that will be covered in this thesis is the proposition of online algorithm for state-space models. Applications will mostly cover environmental sciences, such as pollution tracking, and medical imagery, in collaboration with F. Rousseau (LATIM) [link].

Candidate profile

We are looking for a motivated and talented student holding a Master degree with:

– background in statistics, machine learning, signal processing and applied mathematics
– experience in programming, preferably in Python, and deep learning backends such as TensorFlow, JAX or Torch.

Details

A fully funded PhD position (three-year contract) is available from September/October 2020. This PhD funding will be a main component of an on-going project *DynaLearn*, funded by Labex Cominlabs, and involving other research labs in Brest (IMT Atlantique with LaTIM and Lab-STICC) and Rennes (LETG). It is expected that the PhD candidate will be one the cornerstone of this project. Travel will be planned between the 3 sites. The PhD program will mainly take place at the Université Bretagne Sud located at Campus Tohannic in Vannes [link]. The student will enjoy an international and creative environment where research seminars and reading groups take place very often at both the LMBA UMR CNRS 6205 and IRISA UMR CNRS 6074.

The student will be supervised by:

– Lucas Drumetz [link]: `lucas.drumetz@imt-atlantique.fr`
– Nicolas Courty [link]: `nicolas.courty@univ-ubs.fr`
– François Septier [link]: `francois.septier@univ-ubs.fr`

During his/her PhD, the candidate is expected to target top tier machine learning conferences such as ICML, NeurIPS, AISTATS, and build a solid experience at the crossroad between computational statistics and machine learning.

The candidate is requested to firstly send us a CV and a motivation letter to apply for this position.

References

[1] C. P. Robert and G. Casella, *Monte Carlo statistical methods.*   Springer, 2004.
[2] F. Septier and G. W. Peters, "Langevin and Hamiltonian Based Sequential MCMC for Efficient Bayesian Filtering in High-Dimensional Spaces," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 312–327, Mar. 2016.
[3] D. Prangle, "Distilling importance sampling," *arXiv.org*, Oct. 2019.
[4] A. Thin, N. Kotelevskii, J.-S. Denain, L. Grinsztajn, A. Durmus, M. Panov, and E. Moulines, "MetFlow: A New Efficient Method for Bridging the Gap between Markov Chain Monte Carlo and Variational Inference," *arXiv.org*, Feb. 2020.
[5] D. Rezende and S. Mohamed, "Variational Inference with Normalizing Flows," in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, Eds.   Lille, France: PMLR, July 2015, pp. 1530–1538.
[6] F. Santambrogio, "Optimal transport for applied mathematicians."
[7] G. Peyré, M. Cuturi, *et al.*, "Computational optimal transport," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.
[8] T. Vayer, R. Flamary, R. Tavenard, L. Chapel, and N. Courty, "Sliced Gromov-Wasserstein," in *NeurIPS 2019 - Thirty-third Conference on Neural Information Processing Systems*, vol. 32, Vancouver, Canada, Dec. 2019.
[9] K. Fatras, Y. Zine, R. Flamary, R. Gribonval, and N. Courty, "Learning with minibatch Wasserstein : asymptotic and gradient properties," in *AISTATS 2020 - 23nd International Conference on Artificial Intelligence and Statistics*, ser. PMLR, vol. volume 108, Palermo, Italy, June 2020, pp. 1–20.
[10] T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty, "Optimal Transport for structured data with application on graphs," in *ICML 2019 - 36th International Conference on Machine Learning*, Long Beach, United States, June 2019, pp. 1–16.
[11] L. Ambrogioni, U. Güçlü, Y. Güçlütürk, and M. van Gerven, "Wasserstein variational gradient descent: From semi-discrete optimal transport to ensemble variational inference," *ArXiv*, vol. abs/1811.02827, 2018.