Simulation-based Bayesian inference and deep learning for stochastic modelling

In this project we construct new methods to produce statistical inference for parameters in stochastic models. We wish to exploit the expressive power of *deep learning* (DL), to serve Bayesian inference with *plug-and-play* simulation-based approaches. Several research questions can be formulated to marry DL with simulation-based inference, and here follow a few possibilities.

Simulation-based inference is about easing inference for complex models when standard methodologies cannot be applied, due to analytic or computational difficulties. For a recent, though incomplete, review see [5]. The core of the methodology is to only require forward-simulations from the computer model, rather than model-specific analytic calculation, thus offering a flexible approach. Of course, this is not a silver bullet, and methods should be cleverly researched and designed to produce accurate results within reasonable computational time.

Of specific interest is to produce inference for state-space models (also known as hidden Markov models, [7]) and stochastic differential equations (SDEs, [3]). These models are very widely used in practical applications, however standard inference tools require a high level of technical expertise and are difficult to generalize. A strategy using simulation-based inference should only require forward-simulations from the computer model, thus offering a flexible approach.

An important class of simulation-based inference methods is *approximate Bayesian computation* (ABC, [1,2]). Briefly, ABC methods produce approximations of the parameters' posterior distribution by accepting those computer simulations generating synthetic data that are "close" to observed data. A long standing problem is that, to achieve satisfactory accuracy, ABC requires informative *summary statistics* of synthetic and observed data. One of the possible goals could be to automatically return the required summary statistics via ad-hoc deep neuronal networks. A preliminary successful study is [4]. Additionally we can also construct efficient ways to train the neural networks. Variational autoencoders and stochastic recurrent neuronal networks are likely to be useful here. Further research questions can be formulated for other simulations-based engines, beyond ABC, see below.

It would be of interest to construct deep neuronal networks to ease inference for other types of simulations-based strategies, such as "synthetic likelihoods" [8] or the very computationally expensive class of particle MCMC methods (pMCMC, [6]), where the likelihood is approximated using sequential Monte Carlo approaches. The latter is especially suited for time series data.

All the methods discussed above require consideration of Monte Carlo statistical methods.
The PhD student can decide on which aspects and methods to focus, in agreement with the supervisor.

While the PhD project is of methodological nature, it is of course of interest to use the methodology developed during the project to tackle interesting applied problems. The range of possible applications is endless, and we may consider several challenging case studies, such as (a) models for "large" epidemics, typically consisting of thousands of new cases; (b) smart-city data such as precipitation, pollution data, temperature and humidity. A key challenge here is the development of inference methodology that allows the fitting of partially observed Markov process models (driven by SDEs) to data arriving in near real time. (c) protein folding data.

References

[1] Marin, J. M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, *22*(6), 1167-1180.
[2] Sisson, S. A., & Fan, Y. (2010). Likelihood-free Markov chain Monte Carlo. *arXiv preprint arXiv:1001.2058*.
[3] Fuchs, C. (2013). *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer Science & Business Media.
[4] Wiqvist, S and Mattei, P-A, Picchini, U. and Frellsen, J. (2019). Partially Exchangeable Networks and Architectures for Learning Summary Statistics in Approximate Bayesian Computation. Proceedings of the 36th International Conference on Machine Learning, PMLR 97:6798-6807, 2019.
[5] Cranmer, K., Brehmer, J., & Louppe, G. (2019). The frontier of simulation-based inference. *arXiv preprint arXiv:1911.01429*.
[6] Dahlin, J., & Schön, T. B. (2015). Getting started with particle Metropolis-Hastings for inference in nonlinear dynamical models. *arXiv preprint arXiv:1511.01707*.
[7] Simo Särkkä (2013). Bayesian Filtering and Smoothing. *Cambridge University Press*. Available from Cambridge University Press, PDF available at https://users.aalto.fi/~ssarkka/pub/cup_book_online_20131111.pdf
[8] Leah et al. https://eprints.qut.edu.au/92795/74/92795.pdf