Deep generative models for Monte Carlo sampling

– Ph.D. position in machine learning & statistical signal/image processing –

Nicolas DOBIGEON[1] and Pierre CHAINAIS[2]

[1]University of Toulouse, IRIT/INP-ENSEEIHT, 31000 Toulouse, France

[2]University of Lille, CRIStAL/Centrale Lille, 59000 Lille, France

nicolas.dobigeon@enseeiht.fr, pierre.chainais@ec-lille.fr

Abstract

Numerous machine learning and signal/image processing tasks can be formulated as statistical inference problems. As an archetypal example, recommendation systems rely on the completion of partially observed user/item matrix, which can be conducted via the joint estimation of latent factors and activation coefficients [1, 2]. Similarly, ubiquitous signal/image processing tasks are usually formulated as the estimation of latent (i.e., unobserved) objects or features, whether for low-level processing (e.g., denoising, deconvolution, restoration) or for high-level analysis (e.g., classification, segmentation, feature extraction) [3]. More formally, the object θ to be inferred is usually defined as the solution of a variational or stochastic optimization problem. In particular, within a Bayesian framework, this solution $\hat{\theta}$ is defined as the minimizer of a cost function, referred to as the posterior loss and defined as

$$\hat{\theta} \in \operatorname*{argmin}_{\delta} \mathbb{E}\left[L(\delta, \theta)\big|x\right] \quad \text{with} \quad \mathbb{E}\left[L(\delta, \theta)\big|x\right] = \int L(\delta, \theta) p(\theta|x) d\theta \tag{1}$$

where x denotes the set of available data modeled as the realization of a random variable X fully characterized by the likelihood function $p(x|\theta)$, $p(\theta|x)$ is the posterior distribution related to the likelihood function $p(x|\theta)$ and prior distribution $p(\theta)$ thanks to the Bayes formula and $L(\cdot, \cdot)$ is a given loss function. When this function is chosen as quadratic, i.e., $L(y, z) \triangleq \|y - z\|_2^2$, the Bayesian estimator $\hat{\theta}$ is known to be the posterior mean $\hat{\theta}_{\mathrm{MMSE}} = \mathbb{E}[\theta|x]$ which minimizes the mean square error (MSE) [4]. In most real-world applicative contexts, computing such integrals is not straightforward. One alternative lies in making use of Monte Carlo integration, which consists in approximating any expectation of the form

$$\mathbb{E}[h(Z)|Z \sim f(z)] = \int h(z) f(z) dz \tag{2}$$

by the empirical average

$$\bar{h} = \frac{1}{n} \sum_{i=1}^{n} h(z_i) \tag{3}$$

where $\{z_1, \ldots, z_n\}$ is a sample drawn from the distribution $f(z)$ [5]. Obviously, when dealing with Bayesian inference problems, the distribution $f(z)$ of interest is chosen as the targeted posterior distribution $p(\theta|x)$. This so-called Monte Carlo integration requires the availability of efficient algorithmic schemes able to generate samples from a desired distribution. A huge literature dedicated to random variable generation has proposed various Monte Carlo algorithms. For instance, Markov chain Monte Carlo (MCMC) methods, whose particular instances are the famous Gibbs sampler and Metropolis-Hastings algorithm, define a wide class of algorithms which allow a Markov chain $\{z_1, \ldots, z_n\}$ to be generated with stationary distribution f [6]. Despite their seemingly simplicity and genericity, MCMC algorithms may be computationally inefficient for large-scale and/or highly structured problems, i.e., when the space where z lives is of high dimension or when the distribution $f(z)$ is highly nonlinear or non-standard.

Misspecification or underspecification of the probabilistic model describing the data set and how it relates to the parameters of interest impair most of the statistical inference methods. Indeed, this description, generally of the form of a likelihood function, is a key quantity in basic statistical techniques, such as maximum likelihood and Bayesian estimations. To overcome these issues, approximate Bayesian computation1 (ABC) offers an opportunistic alternative [7]. ABC defines a wide class of methods targeting the posterior distribution, yet bypassing the evaluation of the likelihood function. In its simplest form, known as the

ABC rejection algorithm, only sampling under the prior distribution and the statistical model is required. An important advantage of ABC methods is their solid and well-documented properties, allowing the resulting approximation to be properly quantified.

In this project, the likelihood function is assumed to be implicitly described through large training datasets, from which deep architectures can be deployed. In particular, contrary to discriminative approaches, this project will embed deep generative models, such as variational autoencoders [8], within ABC or splitting-based Monte Carlo methods [9]. Contrary to most works of the literature which have considered these generative learning techniques within a Bayesian framework to face the difficulty of designing appropriate prior distributions, this project takes a different route by leveraging their flexibility in order to describe the observational space.

References

[1] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[2] P. Melville and V. Sindhwani, "Recommender systems," in *Encyclopedia of machine learning*. Springer, 2011, pp. 829–838.

[3] J. Idier, Ed., *Bayesian Approach to Inverse Problems*, ser. Digital Signal and Image Processing. Hoboken, NJ: Wiley-ISTE, 2008.

[4] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. London: Chapman & Hall, 1995.

[5] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer, 2004.

[6] W. Gilks, S. Richardson, and D. Spiegehalter, *Markov Chain Monte Carlo in Practice*. London, UK: Chapman & Hall, 1999.

[7] S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of approximate Bayesian computation*. Chapman and Hall/CRC, 2018.

[8] P. K. Diederik, M. Welling *et al.*, "Auto-encoding variational bayes," in *Proc. Int. Conf. Learning Representations (ICLR)*, vol. 1, 2014.

[9] M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented gibbs sampler – application to large-scale inference problems," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1648–1661, 2019.

Keywords

Machine learning, deep learning, Bayesian inference, Monte Carlo algorithms.

Requirements

The knowledge needed for this work includes a strong background in machine learning, probability, statistics, and/or stochastic simulation. Experience in deep learning will be appreciated.

Scientific environment

Fully funded by the Artificial and Natural Intelligence Toulouse Institute (ANITI), this Ph.D. work will be conducted within the Research Chair lead by

- Nicolas Dobigeon, Professor within the Signal & Communications group at IRIT laboratory (UMR CNRS 5505, Toulouse)

and in collaboration with

- Pierre Chainais, Professor within the SIGMA group at CRIStAL laboratory (UMR CNRS 9189, Lille).

The Ph.D. student will be mainly hosted by IRIT, a joint laboratory of CNRS and University of Toulouse, and located in the INP-ENSEEIHT engineering school, in the city center (see map). The SC group at IRIT has strong expertise in statistical signal/image processing and machine learning, with particular interests in inverse problems with application to remote sensing and biomedical imaging.

Short- and middle-term visits in the SIGMA group at CRIStAL will be planned along the 3-year Ph.D. thesis.