Paris, novembre 2021

Offre de stage

Sujet : Asymptotic analysis of adaptive importance sampling

Possibilité de poursuivre sur une thèse

La Chaire Data Science and Artificial Intelligence for Digitized Industry and Services

Portée par Florence d'Alché-Buc, enseignante-chercheur dans le département Image, Données, Signal de Télécom ParisTech, la chaire DSAI réunit cinq partenaires industriels : Airbus Defence & Space, Engie, Idemia, Safran et Valeo Finance. Son objectif général est de développer, en liaison étroite entre les Parties, une formation et une recherche de niveau international.

Ses quatre principaux axes de recherche sont :

1. Analyse et prévision de séries temporelles (Predictive Analytics on Time Series) ;

2. Exploitation de données hétérogènes, massives et partiellement étiquetées (Exploiting Large Scale and Heterogeneous, Partially Labelled Data) ;

3. Apprentissage pour une prise de décision robuste et fiable (Learning for Trusted and Robust Decision) ;

4. Apprentissage dans un environnement dynamique (Learning through Interactions with a Changing Environment).

Internship proposal :

Asymptotic analysis of adaptive importance sampling

Tutors : Pascal Bianchi, Prof., Telecom Paris, Palaiseau, France, email : prenom.nom@telecom-paris.fr
François Portier, ENSAI, Rennes.

Context : The Monte Carlo simulation framework has become fruitful for exploring probability density functions, especially when the ambient space has a large dimension. Domains of application include for instance computational physics, Bayesian modeling and optimization. Among the most popular Monte Carlo approaches, there are Markov Chains Monte Carlo, sequential Monte Carlo and adaptive importance sampling (AIS). Reference textbooks includes Evans and Swartz (2000), Robert and Casella (2004), Del Moral (2013), Owen (2013). The proposed subject is part of the AIS methodology.
The aim is to iteratively generate a sequence of probability density functions $(q_k)_{k\geq 0}$ on $\mathbb{R}^d \to \mathbb{R}$, which converges, in some sense, to a certain probability density function f, called the *target*. From the algorithmic point of view, the later update rule involves two steps. First, a sample X_k is drawn from the current distribution q_k. Second, the density q_{k+1} is settled on using the newly generated sample. Specifically, a class of AIS algorithms recently studied in Delyon and Portier (2020) is given by :

$$q_{k+1}(x) \propto \sum_{j\leq k} \frac{f(X_j)}{q_j(X_j)} K_h(x - X_j)\,, \tag{1}$$

with a proper renormalization constant, ensuring that q_{k+1} is indeed a probability density function. Here, the function $K_h(\cdot)$, called the *kernel*, plays the role of a mollifier. Otherwise stated, q_k is a smoothed version of the empirical distribution of the sequence $(X_j)_{j\leq k}$, with proper weights. In addition, the kernel K_h depends on a parameter $h > 0$, called the *bandwidth*, acting as a smoothing parameter : a small h results in K_h being peaky at zero (for example, think of a Gaussian distribution with small variance h), whereas a large h results in K_h having a broad support and slow variations. The parameter h is either kept constant, or vanishing with the time k at a convenient rate, but needs in any case to be specified. Under suitable assumptions, it is shown that, with probability one, q_k converges uniformly to the target f,
and the convergence rate is characterized.
In this project, we aim at studying a class of importance sampling algorithm, more general than (1), and to propose a general methodology in order to build new algorithms. Indeed, numerical experimentations have evidenced that it is often beneficial to choose weights different from $\frac{f(X_j)}{q_j(X_j)}$, in order to mitigates the fluctuations and increase the practical convergence speed. So far, this claim is mainly empirical, and misses a theoretical justification. The latter is given the framework of *stochastic approximation* algorithms, as initialy introduced by Robbins and Monro (1951). After a straightforward manipulation, algorithms of the form (1) have the abstract iterative form :

$$q_{k+1} = q_k + \gamma_k H(q_k, X_k),$$

where γ_k is a certain *step size* converging to zero, and $H(q_k, X_k)$ is a certain function on $\mathbb{R}^d \to \mathbb{R}$ which depends on the past estimate q_k of the density, and the newly generated sample X_k. Algorithms of this kind are well known in the optimization community when q_k is a finite-dimensional vector : the celebrated Stochastic Gradient Descent (SGD) is the most famous example. The SGD is classically used in order to minimize a certain *objective function* $F(q)$ with respect to some parameter q, in the case where it is not easy to evaluate F or its gradient. The algorithm (1) turns out to be similar to SGD, when the cost function $F(q)$

is the *Kullback-Leibler divergence* between q and the target density f. Here of course, the parameter q is no-longer finite dimensional, but belongs to a certain class of square-integrable functions, called a reproducing kernel Hilbert space (RKHS).
The aim of the project is the following.

— Review the literature on stochastic approximation algorithms in Hilbert spaces.
— Provide a new class of adaptive importance sampling algorithms, grounded in the stochastic optimization literature.
— Establish closed form estimates of the gap between q_k and the target f, in terms of Kullback-Leibler divergence, α-divergence, uniform norm.
— Run numerical experiments sustaining the efficiency of the proposed algorithms.

Candidacy :The candidate should have a strong background in probability theory. Please send your resume, along with your grades, to the email address provided at the top of this document.

Références

[1] Delyon, Bernard and Portier, François, Safe adaptive importance sampling : A mixture approach, Arxiv preprint 2020