Internship subject

Study of optimization problems to improve importance sampling

Location : ISAE-SUPAERO, Complex Systems Engineering Department (Toulouse, France)
Supervision : Emilien Flayac (ISAE-SUPAERO), emilien.flayac@isae.fr and Florian Simatos (ISAE-SUPAERO), florian.simatos@isae.fr

General context

Importance sampling is a classical method of simulation and estimation which consists in approximating a density f on $\mathbb{R}^d$ using an i.i.d. sample $(X_1, \dots, X_n)$ drawn according to an auxiliary distribution g, weighting each sample X_i with an importance weight w_i. In the case of estimating an integral I of the form

$$I = \mathbb{E}_f(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x) f(x) \mathrm{d}x$$

for a given function $\varphi : \mathbb{R}^d \to \mathbb{R}$, the importance sampling estimator is given by

$$\hat{I}_g = \frac{1}{n} \sum_{i=1}^{n} w_i \varphi(X_i) \quad \text{with} \quad w_i = \frac{f(X_i)}{g(X_i)}.$$

It is well known that the choice of g governs the quality of the estimator $\hat{I}_g$, which depending on g can be more or less efficient than the usual Monte-Carlo estimator (corresponding to the case $g = f$). In the case where $\varphi \geq 0$, the optimal auxiliary density is given by $g_{\text{opt}} = \varphi f / I$, which is not usable in practice as the importance weights $w_i = I/\varphi(X_i)$ are then not computable because of the unknown constant I.
A classic method for choosing g is given by the **cross-entropy method**. In this case, we give ourselves a family $\mathcal{G}$ of auxiliary densities, and seek to choose $g \in \mathcal{G}$ that is as close as possible to g_{opt} in the sense of the Kullback–Leibler divergence, i.e., we look for g solution of

$$\arg\min_{g \in \mathcal{G}} D(g_{\text{opt}}, g) \quad \text{with} \quad D(h, g) = \int_{\mathbb{R}^d} h(x) \ln\left(\frac{h(x)}{g(x)}\right) \mathrm{d}x. \tag{CE}$$

Nevertheless, this approach often leads to auxiliary densities g such that the variables $w_i \varphi(X_i)$ have an infinite second moment, i.e., with

$$\mathbb{E}_g(w_i^2 \varphi(X_i)^2) = \int \left(\frac{f}{g}\right)^2 \varphi^2 g = \int \frac{(\varphi f)^2}{g} = \infty,$$

which degrades the speed of convergence of the estimator $\hat{I}_g$.

Internship goals

The main objective of the internship is to formulate and study optimization problems inspired by (CE) that aim to define auxiliary densities close to g_{opt} in the sense of Kullback–Leibler divergence, but under finite moment constraints. For example, consider the following two problems :

$$\arg\min \left\{ D(g_{\text{opt}}, g) : g \in \mathcal{G} \quad \text{and} \quad \int \frac{(\varphi f)^2}{g} \leq K \right\} \tag{CE-1}$$

and

$$\arg\max \left\{ \beta : g \in \mathcal{G}, \int \frac{(\varphi f)^\beta}{g^{\beta - 1}} < \infty \quad \text{and} \quad D(g_{\text{opt}}, g) \leq K \right\}. \tag{CE-2}$$

The first problem (CE-1) aims to select a density g closest to g_{opt}, but under a constraint on the second moment of $w_i\varphi(X_i)$. And the second problem (CE-2) aims to select a density g with as many finite moments as possible for $w_i\varphi(X_i)$, but under a constraint to remain close to g_{opt}.

The internship will include an optimization dimension, since it will involve formulating problems such as (CE-1) and (CE-2) and trying to solve them for different families of auxiliary distributions $\mathcal{G}$. A first difficulty arises from the fact that the moment constraint that appears in both problems is non-linear in the density g, which is the variable of the problem. A second difficulty arises from the fact that, from a general point of view, the choice of a parametric family generates a nonlinear optimization problem for which specific numerical methods are required. In particular, we will focus on the multivariate Gaussian case, which leads to optimization problems on the space of covariance matrices, i.e., symmetric positive-definite matrices. An attempt will be made to generalize to the case of exponential laws, known to be the optimal class of laws in the presence of linear moment constraints [2].

Secondly, the student will mobilize results from **mathematical statistics** to determine the convergence properties of preferential sampling estimators obtained with the auxiliary densities solutions of (CE-1) and (CE-2).

If the finite-dimensional results prove conclusive, we will also try to study this problem in high dimension, i.e., in the asymptotic regime $d \to \infty$. Indeed, recent results have shown that the solution densities of (CE) are optimal in terms of consistency [1], but in practice, these densities generally have no finite second moment and therefore converge slowly. Conversely, forcing densities with a second moment seems like a good idea from the point of view of convergence speed, but may be suboptimal in terms of consistency. The aim is to understand the trade-offs between consistency and convergence speed in the high-dimensional regime.

Référence

[1] Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

[2] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, Hoboken, NJ, 2001.