Internship and PhD proposal

Subject: On importance sampling for probability estimation of high-dimensional rare events with finite intrinsic dimensions

Supervision:	Xiaoyi Mai (Toulouse Mathematics Institute), xiaoyi.mai@math.univ-toulouse.fr, Florian Simatos (ISAE-SUPAERO), florian.simatos@isae.fr
Location:	Toulouse, France
Work program:	The project is intended for 3.5 year project including a Master's internship followed by a 3-year PhD. The internship will typically focus on the first research question outlined below. The internship is meant to start in the first half of 2026 and the PhD in the second half.
Funding:	Applications are being made to acquire the necessary funding.
Application:	If the topic is potentially interesting to you, please contact us for further discussions. If you wish to apply for the internship and/or the PhD, please provide us a CV, a motivation letter and recent transcripts (at least from the previous year). Recommendation letters are appreciated and may be asked for at a later stage.

Context: importance sampling for rare events in high dimensions
Typically the rare event probability estimation aims to assess a small probability

$$p = \mathbb{P}_f(X \in A),$$

where X is a standard d-dimensional Gaussian vector with density $f(x) = (2\pi)^{-1/2} e^{-\|x\|^2/2}$ and $A \subset \mathbb{R}^d$ some measurable set. The Monte–Carlo estimator

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} \xi_A(X_i),$$

where $\xi_A(X)$ is the indicator function of $X \in A$, has high variance due to the low success rate of $\xi_A(X_i)$. Indeed, to bound the relative variance $\mathrm{Var}(\hat{p})/p^2 = (p - p^2)/np^2 < \epsilon$, it requires $n > (1-p)/\epsilon p$, which is prohibitively high at small p.
To improve the sample efficiency of rare event probability estimation, importance sampling (IS) proposes to estimate p by

$$\hat{p}_g = \frac{1}{n} \sum_{k=1}^{n} \ell(Y_i) \xi_A(Y_i),$$

where the Y_i's are i.i.d. drawn according to an *auxiliary distribution* g, and $\ell = f/g$ is the *likelihood ratio*. This estimator is unbiased and so the law of large numbers readily entails the consistency of $\hat{p}_g$. In adaptive IS schemes, the construction of the auxiliary distribution proceeds in two steps:

1. Identification of a "good" target distribution g_{tar} defined from f and A;
2. Estimation of g_{tar} using n_{tar} samples.

ISAE SUPAERO - Institut Supérieur de l'Aéronautique et de l'Espace
10, avenue Edouard-Belin - BP 54032 - 31055 Toulouse CEDEX 4 - FRANCE
Tél : 33 (0)5 61 33 80 80 - Télécopie : 33 (0)5 61 33 83 30 - Site internet : www.isae-supaero.fr

The oracle choice of g_{tar} is the distribution f conditioned on the rare event set A, i.e., $f|_A = f\xi_A/p$. A single sample Y_1 drawn from $f|_A$ suffices to recover perfectly the rare event probability p, since $\ell(Y_1)\xi_A(Y_1) = p$ holds with probability 1. As $f|_A$ is intractable (it involves the unknown rare event probability p that we seek to estimate), approximations of $f|_A$ are often used instead as target distribution. A common choice is $g_{\text{tar}} = N(\mu_A, \Sigma_A)$, the Gaussian distribution with mean μ_A and variance Σ_A that of $f|_A$. Note that using such target distribution still require estimating μ_A, Σ_A related to the rare event A, which is not less challenging than estimating p. In practice, the estimation of μ_A, Σ_A is carried out in an iterative manner, starting on a large set A_0 than A, then on a series of A_t gradually approaching A, while maintaining a significant success rate of ξ_{A_t}.
In high dimensions, the performance of such schemes collapses due to the low estimation quality of the covariance matrix Σ_A given by

$$\hat{\Sigma}_A = \frac{1}{n_{\text{tar}}} \sum_{i=1}^{n_{\text{tar}}} \ell'(Z_i)(Z_i - \hat{\mu}_A)(Z_i - \hat{\mu}_A)^\top \quad \text{with} \quad \ell' = \frac{f|_A}{g'} \quad \text{and the } Z_i\text{'s i.i.d.} \sim g'. \tag{1}$$

To solve this issue, several authors recently proposed to use auxiliary distributions of the form $N(\hat{\mu}_A, \hat{\Sigma}_{\text{proj}})$ with $\hat{\Sigma}_{\text{proj}}$ the "projection" of $\hat{\Sigma}_A$ on a suitable subspace V, corresponding to $\hat{\Sigma}_{\text{proj}}$ given by

$$\hat{\Sigma}_{\text{proj}} = \sum_{k=1}^{\mathfrak{V}} (\lambda_k - 1) v_k v_k^\top + I \tag{2}$$

with $\mathfrak{V} = \dim(V)$, the v_k's an orthonormal basis of V and $\lambda_k = v_k^\top \hat{\Sigma}_A v_k$.

Main research questions

In practice, several choices for the v_k's in (2) have been proposed. For instance, (4) propose to use $\mathfrak{V} = 1$ and $v_1 = \hat{\mu}_A$, while (6) propose a more complex choice involving the eigenvectors of some suitable matrix. However, **a theoretical understanding on why do these schemes work, and what is the influence of the projection directions** v_k is largely missing. The goal of this project is to fill in this research gap.
As the motivation for covariance matrices of the form (2) stems from the implicit assumption that A can be well approximated by a low-dimensional subspace, we propose to work under the assumption that A **has a finite intrinsic dimension**, in the sense that A is of the form

$$A = \{x \in \mathbb{R}^d : \varphi \circ P(x) \geq 0\} \tag{3}$$

for some projector $P : \mathbb{R}^d \to \mathbb{R}^d$ onto a finite dimensional subspace and some measurable function $\varphi : \mathbb{R}^d \to \mathbb{R}$. In this case, the high-dimensional rare event can be reduced to a small dimensional problem in $\text{Im}(P)$ as (3) implies $x \in A \Leftrightarrow P(x) \in A$.
In this framework, we will aim to address the questions outlined below. The end goal would be to study the behavior of the IS estimator $\hat{p}_{\hat{g}_{\text{tar}}}$ using $\hat{g}_{\text{tar}}$ as auxiliary distribution. To do so, we will first try to assess the influence of the proposed assumptions on the performance of the estimator $\hat{p}_{g_{\text{tar}}}$.

Research question 1: can we caracterize the consistency of $\hat{p}_{g_{\text{tar}}}$ depending on the interplay between U and V?

To go from g_{tar} to $\hat{g}_{\text{tar}}$, it will be necessary to understand the quality of the approximation $\hat{\Sigma}_A \approx \Sigma_A$.

Research question 2: can we use the recent results from random matrix theory (3) to control the quality of the approximation $\hat{\Sigma}_A \approx \Sigma_A$?

Research question 3: can we improve the estimation (1) when A has a finite intrinsic dimension?

Combining these results, we will then be able to study the asymptotic behavior of the IS estimator $\hat{p}_{\hat{g}_{\text{tar}}}$.

Technical details

To address the above research questions, we will leverage results from classical probability theory as well as recent results from random matrix theory.

For the first research question, the goal will be to leverage classical results (e.g., (5)) on the sum of i.i.d. random variables in a triangular setting to derive necessary and sufficient conditions for the convergence $\hat{p}_{g_{\text{tar}}}/p \to 1$. Typically, we will have to control the probability $\mathbb{P}_f(\ell(Y) \geq np\varepsilon \mid Y \in A)$: without further assumptions on g and A this is a difficult problem, but here the idea will be to exploit the particular structure (2) on Σ_A and the assumption on finite intrinsic dimension for A. For instance, we expect results to be quite simple when $U = V_\perp$, but we will seek to investigate more general cases.

For the second research question, we note that (2) can be written in the form $\hat{\Sigma}_A = X^\top L X$ which is a classical model in random matrix theory. However, the originality here is that X is light tailed (as a Gaussian vector) while L is heavy tailed (as a matrix of likelihood ratios). Preliminary results (2; 1) suggest that a phrase transition occurs in the polynomial regime $n = d^\kappa$: if $\kappa > \kappa_*$ for some threshold κ_* then the approximation $\hat{\Sigma}_A \approx \Sigma_A$ is accurate, while if $\kappa < \kappa_*$ the approximation is not accurate. Here the goal will be to extend these preliminary results.

Finally, for the third research question, we will try to generalize the approach of (6): there the authors propose to choose the v_k's as eigenvectors of some empirical matrix $\hat{H}$ whose mean $H = \mathbb{E}(\hat{H})$ has desirable theoretical property. Here the goal will be to prove that when A has a finite intrinsic dimension, then $\hat{H}$ is a low-rank matrix. If this holds, this will shed a new light on why the algorithm of (6) is so efficient. It will also open new perspectives to propose new choices for the projection directions v_k.

References

[1] Jason Beh, Jérôme Morio, and Florian Simatos. Phase transition for conditional covariance matrix estimation by importance sampling in high dimension. In preparation.

[2] Jason Beh, Yonatan Shadmi, and Florian Simatos. Insight from the kullback–leibler divergence into adaptive importance sampling schemes for rare event analysis in high dimension. *Ann. Appl. Probab.*, 35(2):1083–1124, 2025.

[3] Tatiana Brailovskaya and Ramon van Handel. Universality and sharp matrix concentration inequalities. *Geom. Funct. Anal.*, 34(6):1734–1838, 2024.

[4] Maxime El Masri, Jérôme Morio, and Florian Simatos. Improvement of the cross-entropy method in high dimension for failure probability estimation through a one-dimensional projection without gradient estimation. *Reliability Engineering & System Safety*, 216:107991, 2021.

[5] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[6] Felipe Uribe, Iason Papaioannou, Youssef M. Marzouk, and Daniel Straub. Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):818–847, 2021.