Variational autoencoder with weighted samples for high-dimensional non-parametric adaptive importance sampling

Julien Demange-Chryst [1,2], François Bachoc [2], Jérôme Morio [1] & Timothé Krauth [1,3]

[1] *ONERA/DTIS, Université de Toulouse, F-31055 Toulouse, France*
{julien.demange-chryst,jerome.morio}@onera.fr

[2] *Institut de Mathématiques de Toulouse, UMR5219 CNRS, 31062 Toulouse, France*
francois.bachoc@math.univ-toulouse.fr

[3] *Zurich University of Applied Sciences, Centre for Aviation, Winterthur, Switzerland*
timothe.krauth@zhaw.ch

Résumé. L'estimation de densité de probabilité avec des échantillons pondérés est une problématique qui suscite un grand intérêt en statistique. En particulier, il s'agit de la base des algorithmes d'échantillonnage préférentiel adaptatif. Classiquement, une distribution cible peut être approchée soit par un modèle non-paramétrique, soit par un modèle paramétrique. Cependant, ces modèles souffrent du fléau de la dimension ou de leur manque de flexibilité. Dans cette contribution, nous proposons d'utiliser comme modèle approchant une distribution paramétrée par un autoencodeur variationnel. Nous étendons le cadre existant au cas d'échantillons pondérés en introduisant une nouvelle fonction objectif. L'expressivité de cette famille la rend très proche d'un modèle non-paramétrique, et malgré le nombre important de paramètres à estimer, cette famille est beaucoup plus robuste face à la dimension que les familles des distributions gaussiennes et des mélanges de gaussiennes. Par ailleurs, afin d'ajouter encore plus de flexibilité au modèle et qu'il soit notamment capable d'approcher des distributions multi-modales, nous utilisons une distribution *a priori* flexible pour les variables latentes de l'autoencodeur variationnel. De plus, nous introduisons une nouvelle procédure de pré-entraînement de l'autoencodeur variationnel afin de déterminer des bons poids d'initialisation des réseaux de neurones dans le but d'empêcher le phénomène de *posterior collapse*. Enfin, nous décrivons comment utiliser la distribution résultante dans un cadre d'échantillonnage préférentiel afin d'estimer une probabilité de défaillance en grande dimension. Nous illustrons l'intérêt de la méthode proposée sur différents problèmes d'estimation d'évènements rares multi-modaux.

Mots-clés. Autoencodeur variationnel, Échantillon pondéré, Échantillonnage préférentiel adaptatif, Grande dimension, Estimation d'évènement rare, Quantification d'incertitudes.

Abstract. Probability density function estimation with weighted samples is a major topic of interest in statistics. In particular, it is the main foundation of all adaptive importance sampling (IS) algorithms. Classically, a target distribution is approximated either by a non-parametric model or within a parametric family. However, these models suffer from the curse of dimensionality or from their lack of flexibility. In this contribution, we suggest to use as the approximating model a distribution parameterised by a variational autoencoder (VAE). We extend the existing framework of VAEs to the case of weighted samples by introducing a new objective function. The flexibility of this family makes it close to a non-parametric model, and despite the very high number of parameters to estimate, this family is much more efficient in high dimension than the classical Gaussian or Gaussian mixture families. Moreover, in order to add flexibility to the model and to be able to learn multimodal distributions, we use a learnable prior distribution for the latent variable. We also introduce a new pre-training procedure for the VAE to find good starting weights of the neural networks to prevent as much as possible the posterior collapse phenomenon to happen. At last, we explicit how to use the resulting distribution in an IS context, and we introduce the proposed procedure in an adaptive IS algorithm to estimate a rare event probability in high dimension on two multimodal problems.

Keywords. Variational autoencoder, Weighted samples, Adaptive importance sampling, High dimension, Rare event estimation, Uncertainty quantification.

1 General context

Many physical systems are schematically described by a relation of the form $Y = \phi(\mathbf{X})$, where the d-dimensional input vector $\mathbf{X}$ is random and where the output Y is determined through the deterministic function ϕ. A common application is the analysis of a black box model: ϕ represents a numerical code and $\mathbf{X}$ can be regarded as the external conditions in which the calculation is done. The complexity of the numerical model ϕ generally makes impossible to study it analytically. Moreover, calls to the code are supposed to be expensive and can therefore only be made in limited number.

2 Rare event estimation

2.1 Problematic

For safety and/or certification reasons, the **reliability analysis** of the system is a crucial step. Without loss of generality, the failure event can be described as a threshold exceedance event $\{\phi(\mathbf{X}) > t\}$ with $t \in \mathbb{R}$. It is generally a rare event and it is essential to estimate accurately its probability. Crude Monte Carlo sampling techniques are not adapted to a such estimation when the failure probability is low because their convergence requires too many calls to the function ϕ. More efficient methods such as subset sampling [2] or importance sampling [1] have been developed and enable to globally master this issue.

2.2 Importance sampling

Importance sampling (IS) is a well-known uncertainty quantification method, classically used as a variance-reduction technique for Monte Carlo integration including rare event estimation

[10], or for generating points from a target probability distribution known up to a constant [7]. The common denominator of every importance sampling procedure is that they all require to estimate a target probability distribution with weighted samples, and obviously, the accuracy of the algorithm depends on the quality of the estimation of the distribution. Moreover, we also need to be able to not only sample from the built auxiliary distribution, but also to have access to its PDF values.

A first way to perform this density estimation is to use non-parametric models, such as kernel smoothing [13]. These models are flexible, but despite some improvements they strongly suffer from the curse of dimensionality since the size of the required sample to have a good approximation of the target distribution exponentially grows with the dimension. Another solution is to use parametric families of distributions, such as the Gaussian [10] or Gaussian mixture ones [5], which are more robust in medium-high dimension. However, they sometimes require some prior knowledge on the target distribution, and their lack of flexibility and the huge number of parameters to estimate can negatively impact the quality of the estimation when the dimension is high.

3 Variational autoencoder for density estimation

In order to combine both flexibility and robustness faced to the dimension, we propose to use as the auxiliary sampling distribution a distribution parameterised by a variational autoencoder [6], whose main principle has been introduced in the last decade. Variational autoencoders are deep generative models for approximating high-dimensional complex distributions of observed data and generating new samples. The specific feature of a variational autoencoder compared to other density estimation methods is that it performs a dimensionality reduction into a lower dimensional latent space in order to facilitate the estimation. Moreover, in opposition to other dimensionality reduction techniques such as principal component analysis [12] or autoencoders [8], variational autoencoders have good generation properties and give explicitly the approximating distribution, allowing to perform Monte Carlo simulations. This tool is now popular in the machine learning community but not so much in uncertainty quantification.

4 Contribution

4.1 Description of the new proposed procedure

In the present communication [4], we extend the existing framework of variational autoencoders to the case of weighted samples by introducing a new objective function. The resulting IS auxiliary distribution is close to an infinite mixture of Gaussian distributions. Then, its flexibility makes it as expressive as a non-parametric model, and despite the very high number of parameters to estimate, it is much more efficient in high dimension than the classical Gaussian or Gaussian mixture families. Moreover, in order to add even more flexibility to the model and to be able to learn multimodal distributions, we consider a learnable prior distribution for the variational autoencoder latent variables. We also introduce a new pre-training procedure for the variational autoencoder to find good starting weights of the neural networks to prevent as much as possible the posterior collapse phenomenon to happen.

At last, we explicit how the resulting distribution can be combined with importance sampling. Indeed, the existing procedure [11] to compute the PDF values of the resulting distribution of a variational autoencoder leads to a biased and non-convergent importance sampling estimator. In order to keep an unbiased and consistent estimator, we introduce a new way to compute the PDF values. Then, we show how to integrate the whole suggested procedure into existing reliability algorithms, such as the cross-entropy algorithm (CE-VAE), for rare event estimation. Finally, we illustrate the practical interests of the previous efforts on two multi-modal rare-event-estimation problems. The code to reproduce the numerical experiments is publicly available at: `https://github.com/Julien6431/Importance-Sampling-VAE.git`.

4.2 Numerical results

Let us consider the analytical problem [3] given for any even dimension $d \geq 2$ by:

$$\forall \mathbf{x} \in \mathbb{R}^d,\ \psi_1(\mathbf{x}) = \min \left\{ \begin{array}{c} \frac{1}{\sqrt{d}} \sum_{i=1}^{d} x_i \\ -\frac{1}{\sqrt{d}} \sum_{i=1}^{d} x_i \\ \frac{1}{\sqrt{d}} \left(\sum_{i=1}^{d/2} x_i - \sum_{i=d/2+1}^{d} x_i \right) \\ \frac{1}{\sqrt{d}} \left(-\sum_{i=1}^{d/2} x_i + \sum_{i=d/2+1}^{d} x_i \right) \end{array} \right\}. \tag{1}$$

The input vector associated to this function is the standard Gaussian distribution in dimension d. The failure threshold is set to $t = 3.5$ such that the theoretical value of the failure probability is $p_t = 9.3 \times 10^{-4}$. This problem is challenging because it has 4 failure modes. We consider here the problem in dimension $d = 100$. We compare our method with the cross-entropy algorithm using a mixture of von Mises–Fisher–Nakagami distributions (CE-vMFNM) as the auxiliary sampling distribution [9]. The numerical results are given in Table 1.

	CE-VAE	CE-vMFNM3	CE-vMFNM4	CE-vMFNM5
N_{tot}	40000	88000	50000	50000
$\widehat{p}_t^{\text{mean}}$	9.310×10^{-4}	1.319×10^{-3}	9.835×10^{-4}	9.315×10^{-4}
Error (%)	5.31%	512.8%	31.3%	7.56%

Table 1: Comparison of the CE-VAE algorithm with the CE-vMNFM algorithm with different numbers of components and the CE-standard VAE algorithm for the four branches problem over $n_{rep} = 100$ repetitions. The number after the "CE-vMFNM" acronym represents the number of components of the mixture given as an algorithm input.

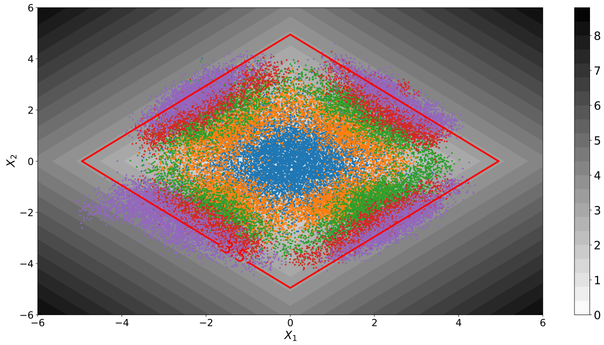

Figure 1: Representation of one realisation of the CE-VAE algorithm on the four branches problem in dimension 2.

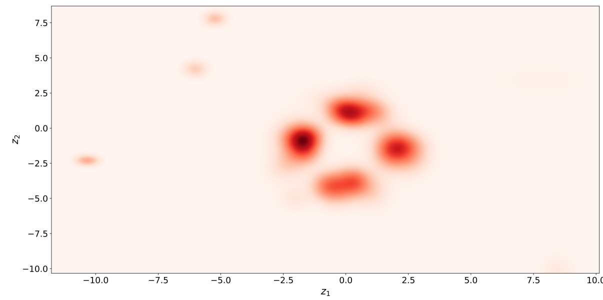

Figure 2: Representation of the two-dimensional PDF of the prior distribution in the latent space associated to the last iteration of one realisation of the CE-VAE algorithm.

We can clearly see that the proposed CE-VAE algorithm provides better performances than the CE-vMFNM algorithm here, and so without any prior knowledge on the number of failure modes. Figure 2 represents the PDF of the prior distribution in the latent space at the last iteration for one execution of the CE-VAE algorithm. One can see that the four failure modes are well represented in it. This is an illustration of the interest of both VampPrior and of the pre-training procedure on a high-dimensional multimodal problem. Therefore when generating new points the four modes will be represented.

References

[1] James Bucklew. *Introduction to rare event simulation.* Springer Science & Business Media, 2004.

[2] Frédéric Cérou, Pierre Del Moral, Teddy Furon, and Arnaud Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–908, 2012.

[3] Marie Chiron, Christian Genest, Jérôme Morio, and Sylvain Dubreuil. Failure probability estimation through high-dimensional elliptical distribution modeling with multiple importance sampling. *Reliability Engineering & System Safety*, 235:109238, 2023.

[4] Julien Demange-Chryst, Francois Bachoc, Jérôme Morio, and Timothé Krauth. Variational autoencoder with weighted samples for high-dimensional non-parametric adaptive importance sampling. *Transactions on Machine Learning Research*, 2024.

[5] Sebastian Geyer, Iason Papaioannou, and Daniel Straub. Cross entropy-based importance sampling using Gaussian densities revisited. *Structural Safety*, 76:15–27, 2019.

[6] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. *Article accepted in the 2nd International Conference on Learning Representations 2014*, 2014.

[7] Jean-Michel Marin, Pierre Pudlo, and Mohammed Sedki. Consistency of adaptive importance sampling and recycling schemes. *Bernoulli*, 25(3):1977 – 1998, 2019.

[8] James L McClelland, David E Rumelhart, PDP Research Group, et al. Parallel distributed processing, volume 2: Explorations in the microstructure of cognition: Psychological and biological models. *MIT press*, 2, 1987.

[9] Iason Papaioannou, Sebastian Geyer, and Daniel Straub. Improved cross entropy-based importance sampling with a flexible mixture model. *Reliability Engineering & System Safety*, 191:106564, 2019.

[10] Reuven Y Rubinstein and Dirk P Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte Carlo simulation, and machine learning*, volume 133. Springer, 2004.

[11] Hechuan Wang, Mónica F Bugallo, and Petar M Djurić. Adaptive importance sampling supported by a variational auto-encoder. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 619–623. IEEE, 2019.

[12] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[13] Ping Zhang. Nonparametric importance sampling. *Journal of the American Statistical Association*, 91(435):1245–1253, 1996.