Visualization of rare event and importance sampling proposal in high-dimensional space

Tianyu Zhang
Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
ty.zhang@connect.ust.hk

Hanzhe CUI
Department of Civil and Environmental Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
hcuiah@connect.ust.hk

Wenxi Qiu
Department of Industrial Engineering and Decision Analytics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
wqiuae@connect.ust.hk

Erjia FU
Department of Industrial Engineering and Decision Analytics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
efuaa@connect.ust.hk

Abstract

Visualizing rare events and importance sampling distributions in high-dimensional spaces is a challenging task. While two-dimensional visualizations provide valuable insights, they become infeasible for problems with three or more dimensions. Researchers in the field of adaptive importance sampling (AIS) have primarily relied on quantitative metrics to assess the quality of importance distributions. In this report, we explore a new perspective by employing dimensionality reduction algorithms to visualize rare events and importance distributions in high-dimensional spaces. By compressing the data into a two-dimensional space, we can intuitively evaluate the performance of importance distributions generated by AIS methods. We investigate three specific methods: Importance Conditional Expectation (ICE) with a Single Gaussian (ICE-SG), ICE with a Gaussian Mixture (ICE-GM), and ICE with the von Mises–Fisher–Nakagami mixture (ICE-vMFNM). We apply these methods to two datasets: a series system of two linear limit state functions and a two-dimensional heat diffusion model. The visualizations obtained through dimensionality reduction provide a qualitative assessment of the importance distributions, complementing the quantitative metrics used in the literature. This new perspective enhances our understanding of the quality of importance distributions and verifies the results of existing quantitative approaches.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

1 Introduction

Reliability analysis is a critical component in the stochastic evaluation of engineering systems, which aims to quantify the probability of a system reaching predefined failure states. Formally, consider a system with n-dimensional uncertain model parameters $\mathbf{x} \in \mathbf{X}$, and let f denotes the failure event for which the limit state function $g(\mathbf{x})$ takes non-positive values. The probability of failure P_f can be defined as:

$$P_f = \int_{\mathbb{R}^n} \mathbb{I}\left(g(\mathbf{x}) \leq 0\right) p(\mathbf{x}) d\mathbf{x} \tag{1}$$

where $p(\mathbf{x})$ is the joint probability distribution function (PDF) for the uncertain variables, $\mathbb{I}(\cdot)$ is the indicator function (1 if the condition is met and 0 otherwise).

The crude Monte Carlo estimation of P_f requires prohibitively many samples to achieve an acceptable accuracy. For example, for P_f around $1 \cdot 10^{-5}$, at least 10^7 samples are required to get an reliable estimation. Therefore, importance sampling is often used to estimate P_f. The importance sampling estimator of P_f is given by:

$$P_f \approx \frac{1}{N} \sum_{i=1}^{N} \frac{p(\mathbf{x}_i)}{q(\mathbf{x}_i)} \mathbb{I}(g(\mathbf{x}_i) \leq 0) \tag{2}$$

where $q(\mathbf{x})$ is the importance distribution, from which $\mathbf{x}_i$ is sampled. According to the importance sampling theory, the optimal proposal distribution is given as follows:

$$q^*(\mathbf{x}) = \frac{\mathbb{I}(g(\mathbf{x}) \leq 0) p(\mathbf{x})}{\int_{\mathbb{R}^n} \mathbb{I}(g(\mathbf{x}) \leq 0) p(\mathbf{x}) d\mathbf{x}} \tag{3}$$

which will lead to 0 estimation variance. However, the optimal proposal distribution is unknown in practice since the denominator is the quantity of interest. Therefore, adaptive importance sampling (AIS) is used to approximate the optimal proposal distribution. In this report, we will consider the improved cross entropy method (ICE) combined with different density models, which is currently one of the most advanced AIS algorithm.

For $n = 2$ problems, both failure domain and $q(x)$ can be directly visualized, providing valuable insights into AIS behavior. However, visualization becomes challenging for problems where $n \geq 3$. In the field of AIS, researchers have traditionally relied solely on quantitative metrics such as relative error and coefficient of variation to evaluate importance distribution quality. This report will explore how dimension reduction algorithms can enhance our ability to assess the quality of importance distributions generated by AIS.

2 Dataset

We will examine two datasets. Denote the labeled dataset as $\{X, y\}$. In both problems, X is sampled from the probability density function $p(\mathbf{x})$, where $p(\mathbf{x})$ represents the standard Gaussian distribution and labels are set as $y = \mathbb{I}(g(\mathbf{x}) \leq 0)$.

2.1 Series system of two linear limit state functions

The limit state function is given as follows,

$$g_1(\mathbf{x}) = \min \left\{ \begin{array}{l} \beta - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \\ \beta + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \end{array} \right\} \tag{4}$$

where we set $n = 100$, leading to a 100-dimensional problem. We choose $\beta = 3.5$ which results in a failure probability of $P_f = 4.65 \cdot 10^{-4}$.

2.2 Two-dimensional heat diffusion model

In this example, we consider a 53-dimensional FEM example, as indicated in Fig. 1. We assume that the temperature field $T(\mathbf{z})$ across this domain is governed by the following partial differential equation:

$$-\nabla \cdot (\kappa(\mathbf{z}) \nabla T(\mathbf{z})) = I_A(\mathbf{z}) Q \tag{5}$$

The boundary conditions include a fixed temperature $T = 0$ on the top boundary and a zero flux condition $\nabla T \cdot \mathbf{n} = 0$ on the remaining boundaries, where $\mathbf{n}$ denotes the normal vector to the boundary. The heat source term $I_A(\mathbf{z})Q$ is localized in a small square domain $A = (0.2, 0.3)$ m $\times$ $(0.2, 0.3)$ m, with a constant value of $Q = 2000$ W/m^3. The indicator function $I_A(\mathbf{z})$ equals 1 when $\mathbf{z}$ belongs to the domain A, and 0 elsewhere. The thermal conductivity $\kappa(\mathbf{z})$ is modeled as a lognormal random field with a mean value of $\mu_\kappa = 1$ W/°C·m and a standard deviation of $\sigma_\kappa = 0.3$ W/°C·m. The random field is characterized by an isotropic squared-exponential autocorrelation function with a correlation length of $\ell = 0.2$ meters. Further details on this problem can be found in the work of Konakli and Sudret [2016]. The quantity of interest is the average temperature within another small square domain $B = (-0.3, -0.2)$ m $\times$ $(-0.3, -0.2)$ m, computed as:

$$T = \frac{1}{|B|} \int_B T(\mathbf{z}) d\mathbf{z} \tag{6}$$

To numerically solve this problem, the random field $\kappa(\mathbf{z})$ is discretized using the expansion optimal linear estimation (EOLE) method with 53 terms. As a result, the model inputs consist of 53 independent standard normal random variables.

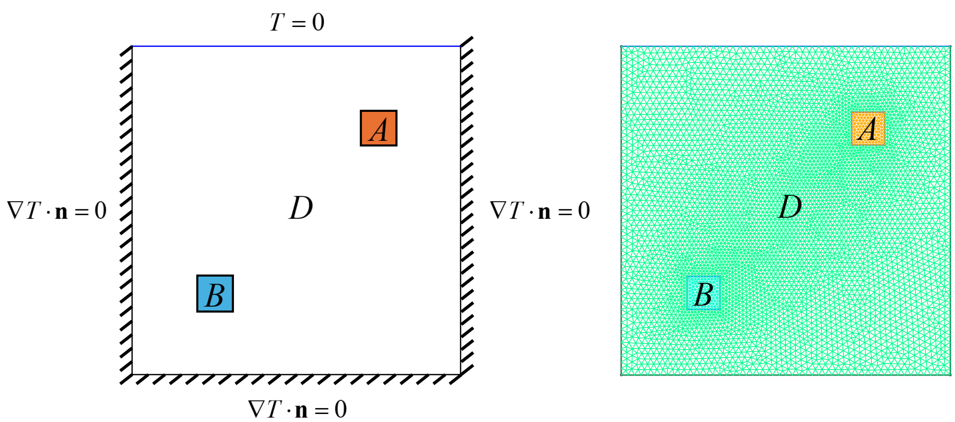

Figure 1: Two-dimensional heat conduction example. Left: domain geometry and boundary conditions; Right: finite element mesh.

The failure event is defined as $f = \{T \geq 7.5°\text{C}\}$ with a reference failure probability of $1.11 \cdot 10^{-3}$. The deterministic solution is obtained using the finite element method with a mesh of 16,000 triangular T3 elements, implemented in MATLAB. The finite element mesh is depicted on the right side of Fig. 1.

3 Experiments settings

Given a dataset X, y, we will apply a dimensionality reduction algorithm to compress the data into a two-dimensional space for visualizing the failure boundary. After fitting the algorithm to the data, we will apply the fitted algorithm to the importance distribution to visualize the samples in the two-dimensional space. From this visualization, we can roughly evaluate the ratio of failure samples generated by the importance distribution, where a higher ratio indicates better quality. We will use this visualization to verify the results of the quantitative metrics used in the paper by Papaioannou et al. [2019].

We will investigate three methods: ICE with a Single Gaussian (ICE-SG), ICE with a Gaussian Mixture (ICE-GM), and ICE with the von Mises–Fisher–Nakagami mixture (ICE-vMFNM).

References

Katerina Konakli and Bruno Sudret. Global sensitivity analysis using low-rank tensor approximations. *Reliability Engineering & System Safety*, 156:64–83, 2016.

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.