Learning Importance Sampling Distributions via Normalizing Flows to Estimate Rare-Event Failure Probability

E.A. Johnson & A. Dasgupta
University of Southern California, Los Angeles, USA.

ABSTRACT

Engineered systems are typically designed to be robust and have very low probabilities of failure. However, developing accurate estimates of these probabilities can be challenging given the complex (and typically nonlinear) nature of the system behavior and the computational cost of simulating a sufficient number of realizations/scenarios to capture the failure modes. Among other approaches, importance sampling can reduce the computational cost and/or the variance in failure probability estimates; however, the optimal importance sampling distribution can only be computed if the failure probability is already known. This work proposes to use normalizing flows (NFs), a type of machine learning model, to learn a near-optimal importance sampling distribution. NFs are generative modeling techniques amenable to exact, but efficient, density evaluation. The approach is first evaluated on a suite of challenging benchmark reliability estimation problems, comparing against two techniques widely adopted for similar tasks: subset simulation and the cross-entropy method; the results show that the proposed approach can be used to estimate rare-event probability in cases that have extremely low failure probabilities on the order of 10^{-7}, high-dimensionality, and multiple failure modes. Finally, the proposed approach is applied to estimate the reliability of a structural mechanics problem.

1 INTRODUCTION

Estimating the probability of a rare event is critical to many safety-critical applications spanning various disciplines of science and engineering. Mathematically, the problem involves estimating the probability of failure P_F, which is given by the following d-fold integral (Beck and Zuev, 2015):

$$\mathrm{P}_F = \int_{\mathrm{g}(\mathbf{x}) \leq 0} p_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int \mathbb{I}\{g(x) \leq 0\}\, p_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{1}$$

where $\mathbf{x} \subseteq \mathbb{R}^d$ is the d-dimensional random vector, representing the input random variables, with joint probability density function (pdf) $p_{\mathbf{X}}(\mathbf{x})$; $g(\mathbf{x})$ is the limit state function (LSF); $\mathbb{I}\{\cdot\}$ is the indicator function; and $F = \{\mathbf{x} \mid g(\mathbf{x}) \leq 0\}$ is the rare event domain. The local maximizers to the constrained optimization problem: $\arg\max_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x})$ such that $g(\mathbf{x}) = 0$, are called the design points or failure modes. Using Monte

Carlo simulation (MCS) to estimate P_F is not bereft of challenges: large sample sizes are necessary to estimate P_F when its value is small, which can render the task computationally intractable. Variance reduction methods attempt to improve the computational efficiency of MCS by increasing the frequency of sampling from F, thereby reducing the required sample size. Importance sampling uses a biasing distribution, also known as the importance sampling distribution, to sample frequently from F. However, the construction of a good ISD relies on knowledge about the number of failure modes: information that is seldom available.

In this work, we propose **REIN** — a new technique for estimating the probability of **r**are **e**vents via **i**mportance sampling, where the ISD is constructed using **n**ormalizing flows (NFs) (Rezende and Mohamed, 2015). Normalizing flows use compositions of invertible neural networks to construct bijective functions that can be used to convert a simple probability density into a more complicated probability density. The task of estimating P_F essentially boils down to training a normalizing flow capable of inducing a good importance sampling distribution. To this end, we propose a loss function to train the normalizing flows that is tailored to rare-event simulation. One advantage of REIN, as we will show, is that it does not rely on or require information about the number of failure modes. We apply REIN to a suite of benchmark reliability estimation problems and a structural reliability problem. The results show that REIN performs well on problems featuring multiple failure modes, high dimensionality, and high nonlinearity.

2 BACKGROUND

2.1 Importance sampling

After introducing the importance sampling distribution $h_{\mathbf{X}}(\mathbf{x})$ into Equation 1, the importance sampling estimator for P_F can be expressed as

$$\mathrm{P}_F^{\mathrm{IS}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\{g(\mathbf{x}^{(i)}) \leq 0\}, \tag{2}$$

where $\mathbf{x}^{(i)}$ is the i^{th} realization of $\mathbf{x}$ drawn from $h_{\mathbf{X}}$. In theory, there exists an optimal importance sampling distribution $h_{\mathbf{X}}^{*}(\mathbf{x})$ for which the variance of the IS estimator is $\mathbb{V}[\mathrm{P}_F^{\mathrm{IS}}] = 0$ irrespective of the sample size N (Tabandeh et al., 2022). However, the normalizing constant of $h_{\mathbf{X}}^{*}(\mathbf{x})$ is P_F, the very quantity that is to be estimated. Hence, sampling from $h_{\mathbf{X}}^{*}(\mathbf{x})$ is impossible. Similar to other parametric importance sampling methods, REIN uses normalizing flows to construct quasi-optimal importance sampling distributions, while utilizing information about the shape of $h_{\mathbf{X}}^{*}(\mathbf{x})$, to efficiently compute P_F.

2.2 Normalizing flows

Normalizing flows (Rezende and Mohamed, 2015) are based on the idea that a simple probability distribution, say p_X, can be transformed into complex probability distributions, like $h_{\mathbf{X}}^{*}$, using a sequence of bijective mappings. Let the *generator* $\mathbf{f}(\mathbf{x};\boldsymbol{\theta})\colon \mathbb{R}^d \to \mathbb{R}^d$ be a bijective and differentiable function parameterized by $\boldsymbol{\theta}$; then, the change of variables formula establishes the following relationship between the *latent* pdf $p_{\mathbf{X}}$ and the *induced* pdf $\mathbf{f}_{\#}p_{\mathbf{X}}$:

$$\mathbf{f}_{\#}p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})\left|\det \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x};\boldsymbol{\theta})\right|^{-1}, \tag{3}$$

where $\det \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x};\boldsymbol{\theta})$ denotes the determinant of the Jacobian matrix of $\mathbf{f}(\cdot;\boldsymbol{\theta})$ (Rezende and Mohamed, 2015). It is difficult to construct generators in high-dimensional spaces that are sufficiently expressive and, therefore, capable of inducing the desired target distribution. Moreover, the computation of $\det \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x};\boldsymbol{\theta})$ and its derivative with respect to $\boldsymbol{\theta}$ must be efficient to enable useful application of probability transformations like Equation 3. Both of these are achieved with normalizing flows by stacking multiple invertible neural networks:

$$\mathbf{f} = \mathbf{f}_{N_\mathrm{f}} \circ \mathbf{f}_{N_\mathrm{f}-1} \circ \ldots \circ \mathbf{f}_1, \tag{4}$$

such that the resultant Jacobian determinant is

$$\det \nabla_\mathbf{x} \mathbf{f}(\mathbf{x}; \boldsymbol{\theta}) = \prod_{k=0}^{N_\mathrm{f}-1} \det \nabla_{\mathbf{x}_{[k]}} \mathbf{f}_{k+1}\left(\mathbf{x}_{[k]}\right), \tag{5}$$

where $\mathbf{x}_{[k+1]} = \mathbf{f}_{k+1}\left(\mathbf{x}_{[k]}\right)$, $\mathbf{x}_{[0]} = \mathbf{x}$ and N_f is the number of flow layers. Each flow layer $\mathbf{f}_k$ is an invertible neural network and $\boldsymbol{\theta}$ now represents the collective parameters of all flow layers.

In this work, we construct invertible neural networks using planar transformations (Rezende and Mohamed, 2015). A planar transformation applies the following invertible transformation to the input $\mathbf{x}$ in flow layer k:

$$\mathbf{f}_k(\mathbf{x}) = \mathbf{x} + \mathbf{u}_k \cdot s\left(\mathbf{w}_k^\mathrm{T}\mathbf{x} + b_k\right) \tag{6}$$

where $\mathbf{u}_k \in \mathbb{R}^d$, $\mathbf{w}_k \in \mathbb{R}^d$ and $b_k \in \mathbb{R}$ are the parameters of $\mathbf{f}_k$, and $\mathrm{s}\colon \mathbb{R} \to \mathbb{R}$ is a nonlinear activation function. with derivative s'. Herein, we use hyperbolic tangent activation. The Jacobian determinant of $\mathbf{f}_k$ is:

$$|\det \nabla_\mathbf{x} \mathbf{f}_k(\mathbf{x})| = \left|1 + \mathbf{u}_k^\mathrm{T} \boldsymbol{\psi}_k(\mathbf{x})\right| \tag{7}$$

where $\boldsymbol{\psi}_k(\mathrm{x}) = \mathrm{s}'\left(\mathbf{w}_k^\mathrm{T}\mathbf{x} + b_k\right)\mathbf{w}_k$. Notably, the number of parameters in every flow layer is $2d + 1$. Thus, the total number of parameters of the normalizing flow also scales linearly with the dimensionality d.

3 PROPOSED METHOD: REIN

3.1 Training the normalizing flow model to construct the importance sampling distribution

When using REIN, the task of estimating P_F boils down to training a NF model with N_f flow layers such that $\mathbf{f}_\# p_\mathbf{X}$ is a good approximation to h^*. We propose to train the NF model using the following loss function:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x} \sim p_\mathbf{X}}[-\log \tilde{g}(\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}); \alpha_t) - \log p_\mathbf{X}(\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}); \alpha_t) - \gamma_t \log |\det \nabla_\mathbf{x} \mathbf{f}(\mathbf{x}; \boldsymbol{\theta})|] \tag{8}$$

where $\tilde{g}(\mathbf{x}; \alpha_t) = \{1 + \exp[\alpha_t g(\mathbf{x})]\}^{-1}$ is the sigmoid function. α_t and γ_t are updated during training as:

$$\alpha_t = \min\{\alpha_\mathrm{end}, 1 + (\alpha_\mathrm{end} - 1)t/T'\} \tag{9}$$

$$\gamma_t = \min\{\gamma_\mathrm{start}, 1 + (\gamma_\mathrm{start} - 1)t/T'\} \tag{10}$$

where the subscript t denotes the t^th training epoch, T is the total number of training epochs, and $T' < T$ is an epoch threshold beyond which the parameters α_t and γ_t are held constant. The gradients of the loss function must be approximated using MCS; we denote N_b to be the MC sample size. It is possible to derive Equation 8 by minimizing the reverse Kullback-Leibler (KL) divergence between $h_\mathbf{X}^*$ and $\mathbf{f}_\# p_\mathbf{X}$, and then replacing the non-differentiable indicator function by $\tilde{g}(\cdot\,; \alpha_t)$ and inflating the target density at epoch t — $h_\mathbf{X}^\dagger \propto \tilde{g}(\cdot\,; \alpha_t) p_\mathbf{X}$ — by γ_t. Inflating the target density increases the chances of sampling from F.

REIN requires as input from the user: the normalizing flow model with number of flow layers N_f, the MC sample size N_b, and the values of α_end and γ_start. The user must also decide on an appropriate stochastic minimization algorithm along with the learning rate in order to train the normalizing flow model.

3.2 Computing the rare event probability using the importance sampling distribution induced by the normalizing flow model

The failure probability P_F can be computed using REIN in two steps:

Step 1. Train a NF model with N_f flow layers. Training essentially involves minimizing the loss function from Equation 6 to determine the parameters $\boldsymbol{\theta}^*$of the normalizing flow model.

Step 2. Estimate P_F using IS with $\mathbf{f}_\# p_\mathbf{X}(\cdot\,; \boldsymbol{\theta}^*)$ as the ISD. The steps for that are:

i. Generate N iid realizations of $\mathbf{x}$ from $p_{\mathbf{X}}$. Let $\mathbf{x}^{(i)}$ denote the i^{th} realization.

ii. Evaluate the LSF for every generator transformed $\mathbf{x}^{(i)}$, i.e., compute $g(\mathbf{f}(\mathbf{x}^{(i)};\boldsymbol{\theta}^*))$.

iii. Compute $\mathbf{f}_{\#}p_{\mathbf{X}}(\mathbf{x}^{(i)};\boldsymbol{\theta}^*)$ using Equation 3.

iv. Estimate P_F as: $\mathrm{P}_F^{\mathrm{REIN}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\{g(\mathbf{f}(\mathbf{x}^{(i)};\boldsymbol{\theta}^*))\} \cdot {p_{\mathbf{X}}(\mathbf{f}(\mathbf{x}^{(i)};\boldsymbol{\theta}^*))}/{\mathbf{f}_{\#}p_{\mathbf{X}}(\mathbf{x}^{(i)};\boldsymbol{\theta}^*)}$

3.3 Statistical properties of the estimator

Since REIN is built on importance sampling, $\mathrm{P}_F^{\mathrm{REIN}}$ is unbiased as long as $\mathbf{f}_{\#}p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}^*) > 0\ \forall\ \mathbf{x} \in F$, which is true since $\mathbf{f}$ is bijective. Moreover, for a perfectly trained generator, which is the case when the Kullback-Liebler divergence between the induced probability density $\mathbf{f}_{\#}p_{\mathbf{X}}$ after training and the target probability density $h_{\mathbf{X}}^{\dagger\dagger} \propto \tilde{g}(\cdot\,;\alpha_{\mathrm{end}})p_{\mathbf{X}}$ equals zero, the variance of the estimator $\mathrm{P}_F^{\mathrm{REIN}}$ is

$$\mathbb{V}[\mathrm{P}_F^{\mathrm{REIN}}] = \frac{1}{N}\left\{\mathrm{P}_{\tilde{F}}\mathbb{E}\left[\frac{\mathbb{I}\{g(\mathbf{x})\leq 0\}}{\tilde{g}(\mathbf{x};\alpha_{\mathrm{end}})}\right] - \mathrm{P}_F^2\right\}, \tag{11}$$

where $\mathrm{P}_{\tilde{F}} = \int \tilde{g}(\mathbf{x};\alpha_{\mathrm{end}})p_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}$. When $\alpha_{\mathrm{end}} > 0$ and after ignoring some higher-order terms, Equation 11 simplifies to

$$\mathbb{V}[\mathrm{P}_F^{\mathrm{REIN}}] = \frac{1}{N}\left\{\mathrm{P}_{\tilde{F}}\mathrm{P}_F - P_{\tilde{F}}\mathbb{E}\left[\mathbb{I}\{g(\mathbf{x}) \leq 0\}\, e^{\alpha_{\mathrm{end}}g(\mathbf{x})}\right] - \mathrm{P}_F^2\right\}. \tag{12}$$

Thus, Equations 11 and 12 imply consistency of the estimator $\mathrm{P}_F^{\mathrm{REIN}}$. Moreover, as $\alpha_{\mathrm{end}} \to \infty$ we have: $\tilde{g}(\mathbf{x};\alpha_{\mathrm{end}}) \to 1$ and $e^{\alpha_{\mathrm{end}}g(\mathbf{x})} \to 0\ \ \forall\, \mathbf{x} \in F$, which ultimately causes $\mathrm{P}_{\tilde{F}} \to \mathrm{P}_F$. Thus, $\mathbb{V}[\mathrm{P}_F^{\mathrm{REIN}}] \to 0$ as $\alpha_{\mathrm{end}} \to \infty$ and the generator $\mathbf{f}$ will have induced the optimal importance sampling distribution. However, α_{end} is some finite value; thus, the induced distribution can at best be quasi-optimal.

4 NUMERICAL EXAMPLES

We compare REIN with the improved cross-entropy (iCE-IS) method (Papaioannou et al., 2019); we explore both Gaussian mixture (GM) and von Mises-Fisher-Nakagami mixture (vMFNM) models as candidate importance sampling distributions. Unless otherwise mentioned, we use a sample size of 5000 for each step of the optimization procedure that determines the parameters of the mixture densities in iCE-IS and set the target coefficient of variation of the importance weights to 1.5. Moreover, we choose the optimal number of mixture components after monitoring the corresponding estimator's performance. We also compare REIN with subset simulation (SS) (Au and Beck, 2001), which remains widely popular because it is not affected by the curse of dimensionality. For SS, we use standard normal proposal distributions conditioned on the current state. The conditional failure probability is chosen to be 0.1. Moreover, the sample size at each level is chosen such that the total number of LSF evaluations is similar to the other methods.

When comparing different variance reduction methods, we measure the performance of any estimator using the normalized root mean square error (nRMSE); a lower value of nRMSE indicates better performance. Similarly, nRMSE $\times \sqrt{N_{\mathrm{call}}}$ is a combined measure for the coefficient of performance such that a low value indicates high efficiency. We evaluate nRMSE from 100 independent runs and record the total number N_{call} of LSF evaluations in each case and report the average. For the IS-based methods, N_{call} includes N; we use $N = 5000$ for the benchmark problems. Wherever necessary, we estimate the reference value of P_F using MCS with sample size 10^9.

We implement REIN using PyTorch and train the NF models using the Adam algorithm. We start training the NF model with zero weight on the $\log\tilde{g}(\mathbf{f}(\mathbf{x};\boldsymbol{\theta});\alpha_t)$ and $\log\, p_{\mathbf{X}}(\mathbf{f}(\mathbf{x};\boldsymbol{\theta});\alpha_t)$ terms for the first 10 epochs; thus, no LSF evaluations are made during this period.

4.1 Benchmark examples

4.1.1 Two-dimensional limit state function with infinitely many design points

Consider the LSF $g_1 = 5.6^2 - (x_1^2 + x_2^2)$ defined in the standard normal space (Arief et al., 2021). LSF g_1 has $\mathrm{P}_F = 1.55 \times 10^{-7}$ and an infinite number of design points. We use $N_\mathrm{f} = 20$ flow layers, $\alpha_\mathrm{end} = 10$, $\gamma_\mathrm{start} = 20$, $T' = 0.9\,T$, $N_\mathrm{b} = 50$, and choose T to match the number of LSF evaluations made by iCE-IS. We set the learning rate at 0.001. REIN and SS yielded estimators with nRMSE equal to 0.046 and 0.056 with $N_\mathrm{call} = 8.4 \times 10^4$ and 26.4×10^4, respectively. The performance of iCE-IS with mixture models was found to be very sensitive to the number of mixture components used. The nRMSE $\times \sqrt{N_\mathrm{call}}$ of the estimators obtained using iCE-IS with GM models were 57.94, 54.98, 49.13, 21.18, 10.29, 23.89 and 22.65 using 1, 2, 4, 8, 12, 16 and 20 mixture components, respectively; in comparison, nRMSE $\times \sqrt{N_\mathrm{call}}$ is equal to 13.33 for REIN. With a similar number of LSF evaluations as REIN, the nRMSE for the best iCE-IS estimator with GM and vMFNM models were 0.036 and 0.067, respectively, with 12 mixture components in both cases. Figure 1 shows some realizations drawn from importance sampling distributions obtained using REIN, and iCE-IS with GM and vMFNM models. On this benchmark, REIN outperforms SS. Although iCE-IS has the capability of outperforming REIN, choosing the optimal number of mixture components is difficult and bound to increase the computational burden. Thus, REIN has the potential to be a *black-box* toolbox for rare event simulation since it can perform when very little knowledge about the problem is available at the outset.

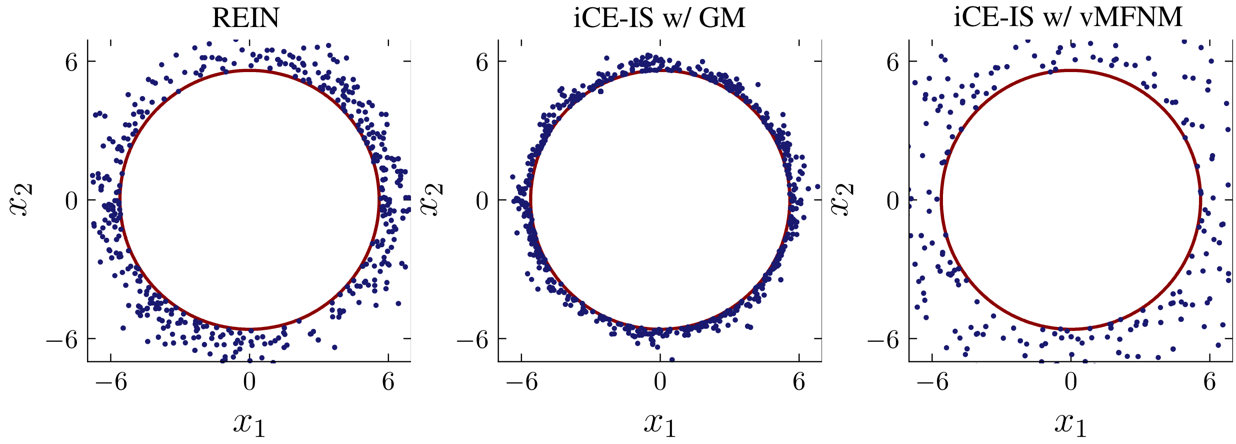

Figure 1: Realizations drawn from the representative ISDs obtained using REIN and iCE-IS

4.1.2 High-dimensional linear limit state function with two design points

LSF g_2 is defined in the standard normal space as the series system of two LSFs:

$$g_2 = \min\{5 \pm d^{-1/2} \textstyle\sum_{i=1}^{d} x_i\}, \tag{13}$$

which has two design points located at equidistant points on two opposite sides of the origin (Papaioannou et al., 2019). The probability of failure P_F is independent of the dimension d and equal to $\mathrm{P}_F = 5.74 \times 10^{-7}$. For this study, we adopt $N_\mathrm{f} = 5$, $\alpha_\mathrm{end} = 10$, $\gamma_\mathrm{start} = 10$, $T = 3000$, $T' = 0.9833T = T - 50$, $N_\mathrm{b} = 100$, and set the learning rate at 0.01. We vary the dimension d between 10 and 500. At dimensions $d > 100$, iCE-IS with a GM model fails to converge; thus, we compare REIN against iCE-IS with vMFNM model using two mixture components, and against SS. Figure 2 shows the performance of REIN on LSF g_2, as d is varied, compared to iCE-IS with a vMFNM model and to SS. $N_\mathrm{call} = 3.05 \times 10^5$ was kept similar for all methods across all dimensions. Again, REIN outperforms SS. However, for dimensions $d \leq 325$, the performance of REIN and iCE-IS with vMFNM distribution is similar, with the latter performing marginally better in some cases. However, note that the number of mixture components to be used in iCE-IS was correctly set equal to the number of design points in this study. Even so, when $d > 325$, the performance of iCE-IS with the vMFNM distribution deteriorates rapidly.

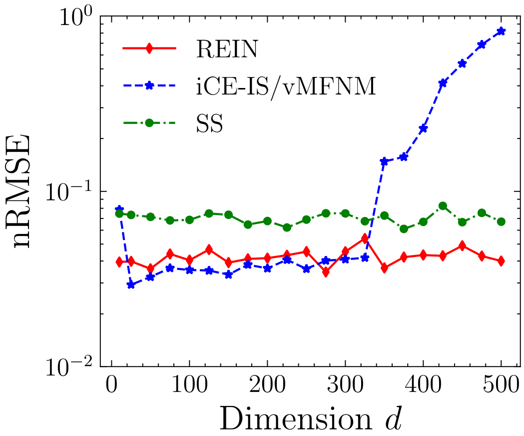

Figure 2: nRMSE of the estimators obtained using REIN, iCE-IS with vMFNM model, and SS for LSF g_2

4.1.3 High dimension highly nonlinear limit state function

Consider the following LSF defined in the standard normal space (Papakonstantinou and Nikbakht, 2020)

$$g_3(\mathbf{x}) = 4 - \frac{1}{\sqrt{d}} \sum_{i=1}^{d} x_i + 2.5\left(x_1 - \sum_{i=2}^{3} x_i\right)^2 + \left(x_4 - \sum_{i=2}^{6} x_i\right)^4 + \left(x_7 - \sum_{i=8}^{9} x_i\right)^8 \quad (14)$$

In this study, we vary the dimension $d \in \{100, 200, 300\}$. We use a normalizing flow model with $N_\mathrm{f} = 75$ flow layers, a learning rate of 0.01, batch size $N_\mathrm{b} = 100$, and train for $T = 3000$ epochs. iCE-IS with vMFNM models failed to converge even after we increased the sample size used in each iteration to 10^5. Table 1 lists the performance of REIN and SS in this example. REIN outperforms SS by a significant margin.

Table 1: Comparison of performance of REIN and SS on the LSF g_3 for varying dimension d (Bold indicates best performance).

d	P_F	REIN		SS	
		nRMSE	N_{call}	nRMSE	N_{call}
100	3.43×10^{-7}	0.120	3.05×10^{-7}	0.269	3.28×10^5
200	3.55×10^{-7}	0.158	3.05×10^{-7}	0.265	3.28×10^5
300	3.71×10^{-7}	0.134	3.05×10^{-7}	0.262	3.28×10^5

4.2 Applications

We consider a high dimensional nonlinear engineering application (Papakonstantinou and Nikbakht, 2020): a thirty-four-story shear frame subjected to a static load F_i at each floor level i, as shown in Figure 3. These lateral loads could represent static equivalent seismic or wind forces. The floor slabs, assumed rigid, are supported on two columns, each of 4 m length, with flexural rigidities $EI_{1,i}$ and $EI_{2,i}$, respectively.

The static loads and flexural rigidities of the columns are assumed to be normally distributed with means 2 kN and 20 MN·m^2, respectively, and coefficients of variation 0.4 and 0.2, respectively. Thus, there are a total of 102 random variables, i.e., $d = 102$, in this application. The LSF is given as:

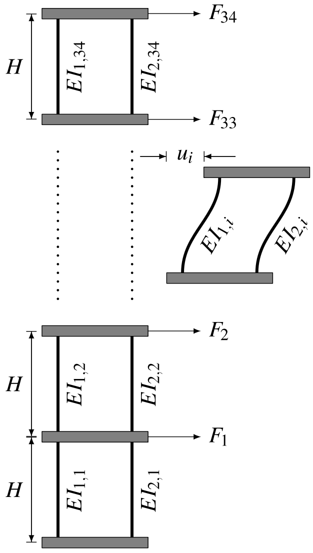

Figure 3: Thirty-four story shear frame subject to static lateral loads at every floor level

$$g = 0.235 - \left|\sum_{i=1}^{34} u_i\right|, \tag{15}$$

where u_i is the i^{th} inter-story drift, which is computed as

$$u_i = {H^3 \sum_{j=1}^{i} F_i} \Big/ {12\left(EI_{1,i} + EI_{2,i}\right)}. \tag{16}$$

The reference value of $\mathrm{P}_F = 2.63 \times 10^{-7}$ is computed using MCS with 10^9 realizations. In this application we use a normalizing flow model with $N_\mathrm{f} = 10$ flow layers that is trained for $\mathrm{T} = 3000$ epochs at a learning rate of 0.001 and batch size $N_\mathrm{b} = 100$. We also use $\alpha_\text{end} = 10$, $\gamma_\text{start} = 4$ and $T' = 0.9T$. Finally, we use $N = 5000$ to compute $\mathrm{P}_F^{\text{REIN}}$. For iCE-IS, we use a sample size of 500 in each iteration. REIN yields an estimator with nRMSE = 0.064 for $N_\text{call} = 3.05 \times 10^5$. In comparison, with a similar number of LSF evaluations, iCE-IS with vMFNM yielded an estimator with nRMSE of 0.464. In the case of SS, the estimator had nRMSE = 0.085, albeit for $N_\text{call} = 3.28 \times 10^5$. REIN outperforms the other variance reduction methods on this high dimensional problem since its estimator achieved the lowest value of nRMSE for a similar number of LSF evaluations.

5 CONCLUSIONS

In this paper, we introduce REIN, which utilizes NFs to construct ISDs that can be used to efficiently estimate rare-event probabilities. Through various examples, we have demonstrated that REIN can be a potential tool for rare-event probability estimation where little to no prior knowledge about the rare-event domain and/or the number of failure modes is known. Further, REIN was found to outperform iCE-IS and SS on all moderate to high-dimensional examples. Future work should explore the effects of various hyperparameters — α_end, γ_start, T, T', N_b and N_f — on REIN.

6 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of this work by the National Science Foundation through award CMMI 16-63667. The first author also acknowledges support from the University of Southern California through the Provost Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or USC. The authors also acknowledge the Center for Advanced Research Computing at the University of Southern California for providing computing resources.

7 REFERENCES

Au SK and Beck JL (2001). "Estimation of small failure probabilities in high dimensions by subset simulation." *Probabilistic Engineering Mechanics,* **16**(4): 263–277.

Arief M, Bai Y, Ding W, He S, Huang Z, Lam H and Zhao D (2021). "Certifiable deep importance sampling for rare-event simulation of black-box systems." *arXiv preprint.* **arXiv**:2111.02204.

Beck JL and Zuev KM (2015). "Rare event simulation." *arXiv preprint.* **arXiv**:1508.05047.

Papaioannou I, Geyer S and Straub D (2019). "Improved cross entropy-based importance sampling with a flexible mixture model." *Reliability Engineering & System Safety* **191**: 106564.

Papakonstantinou KG and Nikbakht H (2020). "Hamiltonian MCMC methods for estimating rare events probabilities in high-dimensional problems." *arXiv preprint.* **arXiv**: 2007.00180.

Rezende DJ and Mohamed S (2015). "Variational inference with normalizing flows." *International Conference on Machine Learning*, 6 July – 11 July, Lille, France, pp. 1530-1538.

Tabandeh A, Gaofeng J and Gardoni P (2022). "A review and assessment of importance sampling methods for reliability analysis." *Structural Safety* **97**: 102216.