Adaptive importance sampling for seismic fragility curve estimation

Clément Gauchy[a,b], Cyril Feau[a], Josselin Garnier[b]

[a] *Université Paris-Saclay, CEA, Service d'Études Mécaniques et Thermiques, 91191, Gif-sur-Yvette, France*
[b] *CMAP, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau Cedex, France*

Abstract

As part of Probabilistic Risk Assessment studies, it is necessary to study the fragility of mechanical and civil engineered structures when subjected to seismic loads. This risk can be measured with fragility curves, which express the probability of failure of the structure conditionally to a seismic intensity measure. The estimation of fragility curves relies on time-consuming numerical simulations, so that careful experimental design is required in order to gain the maximum information on the structure's fragility with a limited number of code evaluations. We propose and implement an active learning methodology based on adaptive importance sampling in order to reduce the variance of the training loss. The efficiency of the proposed method in terms of bias, standard deviation and prediction interval coverage are theoretically and numerically characterized.

Keywords: Computer experiments, probabilistic risk assessment, importance sampling, statistical learning

1. Introduction

The notion of fragility curve was developed in the early 80s in the context of seismic probabilistic risk assesment (SPRA) [1, 2] or performance based earthquake engineering (PBEE) [3]. Fragility curve expresses the probability of

⋆Fully documented templates are available in the elsarticle package on CTAN.

failure of a mechanical structure subjected to a seism conditionally to a specific seismic intensity measure (IM) such as the peak ground acceleration of the seism. Fragility curves are used in several domains: nuclear safety evaluation [4], estimation of the collapse risk of structures in seismic regions [5], design checking process [6]. Nonetheless, the use of fragility curves is not limited to seismic load but is extended to other loading sources such as wind and waves [7]. For complex structures, fragility curve estimation requires a large number of numerical mechanical simulations, involving in most cases non linear computationally expensive calculations. Moreover, they should account for both the uncertainties due to the seismic demand and due to the lack of knowledge on the system itself, respectively called random and epistemic uncertainties [8, 9]. As failure for a typical and reliable mechanical structure is a rare event, the crude Monte Carlo method cannot be applied because it would require too many numerical simulations to produce a sufficiently large number of failed states [10, p.27]. The computational burden motivates the use of parametric models for fragility curves. A typical approach is to assume a lognormal form as first proposed by Kennedy et al. [1] and then widely used (see e.g. [11, 12, 13, 14]). Another way to reduce the computational cost is to fit statistical metamodels to the structure response, such as kriging [15], support vector machines [16], or polynomial chaos expansion [17].

Nonetheless, for the fragility curves to be useful decision-making tools in an industrial context, their estimations must require a number of numerical code calls as small as possible while having theoretical guarantees on the convergence of the estimates and their confidence intervals. Statistical analysis regarding fragility curves estimations - including hypothesis testing and confidence intervals estimation - has been firstly introduced by Shinozuka et al. in [11]. Then, several works have addressed these issues from a non-theoretical point of view (see for e.g. [18, 19, 20, 16]).

In this work, we propose and implement a methodology based on Adaptive Importance Sampling (AIS) [21, 22] in a statistical learning context [23]. Such methods have already been discussed, implemented and tested for proba-

bility estimation of rare event (e.g. failure state) in reliability analysis [24, 25]. We show by asymptotic analysis and numerical simulations that they allow for fast convergence of the estimated fragility curve to the true (unknown) fragility curve and that the uncertainty of the estimation can be rigorously quantified. The uncertainty quantification involves asymptotic confidence intervals and ellipsoids for the quantities of interest and statistical tests to determine whether the asymptotic regime has been reached and whether asymptotic confidence intervals and ellipsoids can be used.

The methodology is applied to different test cases in the application section 5 and compared with more traditional approaches. For the test cases, the stochastic model of modulated and filtered white-noise process defined in [26] is used in order to enrich a set of real ground motion records selected in a database using magnitude and distance criteria. This stochastic model is chosen because it encompasses well both temporal and spectral nonstationarities of real seismic signals and has already been used in [27, 28, 16]. As in [16], 97 acceleration records selected from the European Strong Motion Database [29] in the domain $5.5 < M < 6.5$ and $R < 20$km - where M is the magnitude and R the distance from the epicenter - are considered in order to identify the parameters of the ground-motion model. 10^5 realizations of synthetic signals are then generated. Regarding the parametric approximations of the fragility curves, the current lognormal model is used, taking into account only the variability of the ground motion. Finally, Peak Ground Acceleration (PGA) and the Pseudo-Spectral Acceleration (PSA) are considered to be IMs.

The proposed methodology relies on parametric approximations of fragility curves for any IM of interest. Although the validity of parametric models is both questionable and difficult to assess (see e.g. [13, 27, 30, 20]), some numerical experiments based on the seismic responses of simple mechanical systems - i.e few degrees of freedom systems - suggest that the choice of an appropriate IM makes it possible to reduce the potential biases between reference fragility curves - that can be obtained by massive Monte Carlo simulations - and their parametric approximations. This point will be illustrated in the application

section 5 of this paper. In practice, the selection of an optimal IM is not a trivial matter (see e.g. [31, 32]) and Machine Learning techniques can be used for this purpose (e.g. [16]), knowing that the references [33] and [34] give optimality criteria for selection of such IM.

The paper is organized as follows. In Section 2 the estimation problem is defined for generic parametric fragility curves and any IM. Section 3 is dedicated to AIS algorithm for seismic fragility curve estimation, including theoretical properties and a stopping criterion to check AIS convergence. Section 4 provides the reader with a practical implementation of AIS for seismic fragility curve estimation. Finally, in Section 5, AIS performance is assessed on analytical and industrial test cases of increasing complexity.

2. Seismic fragility curve estimation: a statistical learning framework

We consider the following situation. Let $\mathcal{X}$ be a compact set of $\mathbb{R}$, X a $\mathcal{X}$-valued random variable and $S \in \{0,1\}$ a random label. In seismic probabilistic risk assessment, a structure is subjected to a seismic load, IM is the intensity measure of the seismic load, $X = \log IM$ (we could have $X = \psi(IM)$ where ψ is another increasing function of IM, such as a Box-Cox transform [35]), and S is the indicator variable of the failure of the structure. The pair (X, S) has the probability distribution P over $\mathcal{X} \times \{0,1\}$:

$$P(dx, ds) = \big[\mu(x)\delta_1(ds) + (1-\mu(x))\delta_0(ds)\big]p(x)dx \ , \tag{1}$$

where p is the marginal probability density function (pdf) of X and the fragility curve $\mu(x)$ is the conditional expectation of S (conditional probability of failure or fragility curve):

$$\mu(x) = \mathbb{E}[S|X = x] \ .$$

The aim of the paper is to estimate μ from datapoints $(X_i, S_i)_{i=1}^n$ that may be independent and identically distributed with the distribution P or that may be selected by a more appropriate scheme. There are numerous methods for estimating fragility curves, the interested reader is referred to [36, 37] for a

review of the usual existing methods. It is a classical assumption to use a parametric form for the fragility curve μ to tackle the need for time consuming mechanical simulations, we thus consider the space of functions $\mathcal{F} = \{f_\theta, \theta \in \Theta\}$, where $x \mapsto f_\theta(x)$ is a function from $\mathbb{R}$ to $[0,1]$ for any θ and $\Theta \subset \mathbb{R}^m$. The goal is to minimize the quadratic risk:

$$g(\theta) = \mathbb{E}[(\mu(X) - f_\theta(X))^2] \,,$$

in order to find (provided it exists and is unique):

$$\theta_* = \underset{\theta \in \Theta}{\operatorname{argmin}}\, g(\theta) \ .$$

Unfortunately, the observable data are $(X_i, S_i)_{i=1}^n$, we do not observe directly $\mu(X_i)$. But considering that:

$$\mathbb{E}[(S - f_\theta(X))^2] = \mathbb{E}[\mu(X)(1-\mu(X))] + \mathbb{E}[(\mu(X) - f_\theta(X))^2] \,, \tag{2}$$

we can observe that the minimization of $\mathbb{E}[(S - f_\theta(X))^2]$ is equivalent to the minimization of $\mathbb{E}[(\mu(X) - f_\theta(X))^2]$. Hence, we will consider the quadratic risk

$$r(\theta) = \mathbb{E}[(S - f_\theta(X))^2] \ .$$

In the context of classical learning, when we observe n datapoints $(X_i, S_i)_{i=1}^n$ drawn independently from the probability distribution $P(dx, ds)$ over $\mathcal{X} \times \{0,1\}$, the expectation can be approximated by the empirical mean:

$$\widehat{R}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} (S_i - f_\theta(X_i))^2 \ . \tag{3}$$

The corresponding passive estimator (the term passive is used to highlight the absence of any particular sampling strategy) is then:

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \widehat{R}_n(\theta). \tag{4}$$

As part of industrial applications, the label S_i is obtained by a computationally
80 expensive numerical simulation of a mechanical structure subjected to a random seismic load and X_i is the intensity measure of the seismic load. The seismic

load is an inexpensive artificial seismic signal to generate while the numerical mechanical simulations are computationally expensive.

Conversely to classical learning, active learning aims at selecting the most useful numerical experiments to be carried out in order to form the learning set. In the passive strategy, the datapoints X_i are sampled from the original probability distribution with density p drawn from the artificial realizations of the stochastic ground-motion model. Our active learning strategy, called Adaptative Importance Sampling (AIS), consists to draw the datapoints X_i from an instrumental probability distribution with pdf q that is chosen in an adaptive way. In our context, it is straightforward to use a rejection method to generate seismic loads with a desired intensity measure distribution.

The main heuristic of this procedure is to reduce the variance implied by the empirical approximation of the quadratic risk $r(\theta)$. Importance Sampling is a classical variance reduction technique for Monte Carlo estimation used in structural reliability [38, 25]. If the $(X_i)_{i=1}^n$ are sampled with the pdf q and $(S_i)_{i=1}^n$ are the labels obtained from n calls to the mechanical model of the structure subjected to the seismic loads with intensity measures $(X_i)_{i=1}^n$, then the importance sampling estimator of the empirical quadratic risk is:

$$\widehat{R}_{n,q}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)}{q(X_i)}(S_i - f_\theta(X_i))^2. \tag{5}$$

In the rest of the paper, we will denote by $r(\theta) = \mathbb{E}_{(X,S)\sim P}[\ell_\theta(X,S)]$ with $\theta \mapsto \ell_\theta(x,s)$ a positive loss function for the sake of generalization. For the numerical applications, only the case of the quadratic loss $l_\theta(x,s) = (s - f_\theta(x))^2$ will be considered. The next section will focus on choosing an optimal density q in an adaptive way.

3. Adaptive importance sampling (AIS)

3.1. Principle

The heuristic used to find a good instrumental probability distribution family is presented in [39]. The first idea would be to minimize the variance of the

importance sampling risk estimator (5):

$$\mathrm{Var}\big(\widehat{R}_{n,q}(\theta)\big) = \frac{1}{n}\Big\{ \iint_{\mathcal{X}\times\{0,1\}} \frac{p(x)^2}{q(x)} \ell_\theta(x,s)^2 P(dx,ds) - r(\theta)^2 \Big\} \ , \qquad (6)$$

with respect to q within the set of all pdfs. If we denote by $\tilde{\ell}_\theta(x) = \mathbb{E}[\ell_\theta(X,S)^2 | X = x]$ the loss averaged on S:

$$\tilde{\ell}_\theta(x) = \mu(x)\ell_\theta(x,1)^2 + (1-\mu(x))\ell_\theta(x,0)^2 \ , \qquad (7)$$

the variance of the importance sampling risk estimator (5) can be expressed as

$$\mathrm{Var}\big(\widehat{R}_{n,q}(\theta)\big) = \frac{1}{n}\Big\{ \int_{\mathcal{X}} \frac{p(x)^2}{q(x)} \tilde{\ell}_\theta(x) dx - r(\theta)^2 \Big\} \ ,$$

and we look for

$$q_\theta^* = \operatorname*{argmin}_q \int_{\mathcal{X}} \frac{p(x)^2}{q(x)} \tilde{\ell}_\theta(x) dx \ . \qquad (8)$$

Using Lagrange multipliers, we can solve the optimization problem and we can find that the optimal sampling pdf is of the form

$$q_\theta^*(x) \propto \sqrt{\tilde{\ell}_\theta(x)} p(x) \ ,$$

which depends on μ because $\tilde{\ell}_\theta$ depends on μ [Here and below $\propto$ means equality up to a multiplicative constant]. Hence an approximation step is made by replacing μ by f_θ in (7):

$$\tilde{\ell}_\theta(x) \approx f_\theta(x)\ell_\theta(x,1)^2 + (1-f_\theta(x))\ell_\theta(x,0)^2 \ . \qquad (9)$$

Hence the instrumental density becomes:

$$q_\theta(x) \propto p(x)\sqrt{f_\theta(x)\ell_\theta(x,1)^2 + (1-f_\theta(x))\ell_\theta(x,0)^2} \ . \qquad (10)$$

Note that the instrumental distribution depends on θ, the parameter we aim to estimate. The procedure for computing the AIS estimator $\tilde{\theta}_n$ is described in Algorithm 1. Note also that the algorithm needs to start from a certain parameter value $\tilde{\theta}_0$. We will discuss this issue in section 4.2.

Algorithm 1 Adaptive Importance Sampling

1. Initialization: $\tilde{\theta}_0$
2. for $i = 1, \dots, n$:
   (a) Draw X_i from the distribution with pdf $q_{\tilde{\theta}_{i-1}}$
   (b) Call the mechanical simulation at point X_i to get label S_i
   (c) Compute

$$\tilde{\theta}_i = \operatorname*{argmin}_{\theta \in \Theta} \tilde{R}_i(\theta), \qquad \tilde{R}_i(\theta) = \frac{1}{i} \sum_{j=1}^{i} \frac{p(X_j)}{q_{\tilde{\theta}_{j-1}}(X_j)} \ell_\theta(X_j, S_j) \tag{11}$$

3.2. Theoretical results

We derive some theoretical properties for the estimator $\tilde{\theta}_n$, consisting in its consistency and asymptotic normality. Detailed proofs of the following theorems are given in the Appendix. In the following, if $\theta \mapsto f(\theta)$ is a function, then $\dot{f}$, resp. $\ddot{f}$, $\dddot{f}$, stands for the gradient, resp. the Hessian matrix, the tensor of the third-order derivatives, of f.

Theorem 1. *[Consistency of $\tilde{\theta}_n$] Assume $\theta_* = \operatorname{argmin}_{\theta \in \Theta} r(\theta)$ exists and is unique. Set $L(x, s) = \sup_{\theta \in \Theta} \ell_\theta(x, s)$. Assume that Θ is a compact set and that:*

$$\iint L(x,s) P(dx, ds) < +\infty,$$

$$\sup_{\theta \in \Theta} \iint \frac{L(x,s)^2 p(x)}{q_\theta(x)} P(dx, ds) < +\infty,$$

$$\forall \theta \neq \theta_*, \iint \ell_\theta(x,s) P(dx,ds) > \iint \ell_{\theta_*}(x,s) P(dx,ds).$$

If for any $(x, s) \in \mathcal{X} \times \{0, 1\}$, $\theta \in \Theta \mapsto \ell_\theta(x, s)$ is continuous, then

$$\tilde{\theta}_n \to \theta_* \text{ in probability}$$

Theorem 2. *[Asymptotic normality of $\tilde{\theta}_n$] Assume that $\theta \mapsto \ell_\theta$ is three times differentiable at θ_* for all x, s and that the matrix $\ddot{r}(\theta_*)$ exists and is nonsingular. Assume that the third-order derivatives of $\theta \mapsto \ell_\theta(x, s)$ are dominated in a*

neighborhood of θ_* by a function that is integrable with respect to P. Assume also that the following conditions are satisfied:

1. *The hypotheses of Theorem 1 are satisfied,*
2. $\exists \eta > 0$ *such that* $\sup_{\theta\in\Theta} \iint ||\frac{p(x)\dot{\ell}_{\theta_*}(x,s)}{q_\theta(x)}||^{2+\eta} P(dx,ds) < +\infty$,
3. $\sup_{\theta\in\Theta} \iint \frac{p(x)||\ddot{\ell}_{\theta_*}(x,s)\ddot{\ell}_{\theta_*}(x,s)^T||}{q_\theta(x)} P(dx,ds) < +\infty$,
4. *there exists a neighborhood $\mathcal{B}$ of θ_* such that* $\forall (x,s) \in \mathcal{X}\times\{0,1\}$, $\sup_{\theta\in\mathcal{B}} \frac{p(x)\|\dddot{\ell}_\theta(x,s)\|}{q_\theta(x)} < +\infty$.

Then the sequence $\sqrt{n}(\tilde{\theta}_n - \theta_*)$ is asymptotically normal with mean zero and covariance matrix

$$G_{\theta_*} = \ddot{r}(\theta_*)^{-1} V_{\theta_*} (\ddot{r}(\theta_*)^{-1})^T \,, \tag{12}$$

where V_{θ_*} is the matrix

$$V_{\theta_*} = \mathbb{E}\left[\frac{p(X)\dot{\ell}_{\theta_*}(X,S)\dot{\ell}_{\theta_*}(X,S)^T}{q_{\theta_*}(X)}\right]. \tag{13}$$

Theorem 3. *[Asymptotic normality of the loss] If the conditions of Theorem 2 are satisfied then the sequence $\sqrt{n}(\tilde{R}_n(\tilde{\theta}_n) - r(\theta_*))$ is asymptotically normal with mean zero and variance T_{θ_*} with*

$$T_{\theta_*} = \mathbb{E}\left[\left(\frac{p(X)}{q_{\theta_*}(X)}\ell_{\theta_*}(X,S) - r(\theta_*)\right)^2\right]. \tag{14}$$

A straightforward corollary of Theorem 2 is that, if G_{θ_*} is nonsingular (which we assume from now on), then for any $\alpha \in (0,1)$:

$$\mathbb{P}\left(n(\tilde{\theta}_n - \theta_*)^T G_{\theta_*}^{-1}(\tilde{\theta}_n - \theta_*) < q^\alpha_{\chi^2(m)}\right) \xrightarrow[n\to+\infty]{} \alpha \,, \tag{15}$$

with $q^\alpha_{\chi^2(m)}$ the α-quantile of the $\chi^2(m)$ distribution. Remark that the matrix G_{θ_*} depends on the unknown parameter θ_*. It is thus possible to use a plug-in estimator:

$$\widehat{G}_n = \widehat{\ddot{r}}_n(\tilde{\theta}_n)^{-1} \widehat{V}_n(\tilde{\theta}_n)(\widehat{\ddot{r}}_n(\tilde{\theta}_n)^{-1})^T \,, \tag{16}$$

with

$$\widehat{\ddot{r}}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{p(X_i)}{q_{\tilde{\theta}_{i-1}}(X_i)} \ddot{\ell}_\theta(X_i,S_i) \,, \tag{17}$$

$$\widehat{V}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(X_i)^2}{q_\theta(X_i) q_{\tilde{\theta}_{i-1}}(X_i)} \dot{\ell}_\theta(X_i, S_i)\dot{\ell}_\theta(X_i, S_i)^T \,, \tag{18}$$

and $\ddot{\ell}_\theta(x,s)$ the Hessian of $\ell_\theta(x,s)$ with respect to variable θ. The following lemma establishes that $\widehat{G}_n^{-1} \to G_{\theta_*}^{-1}$ in probability and it is proved in Appendix E.

Lemma 1. *[Uniform convergence of $\widehat{G}_n^{-1}$] Set $L_{1,k,l}(x,s) = sup_{\theta\in\Theta}\ddot{\ell}_\theta(x,s)_{k,l}$ and $L_{2,k,l}(x,s) = sup_{\theta\in\Theta}\frac{p(x)}{q_\theta(x)}(\dot{\ell}_\theta(x,s)\dot{\ell}_\theta(x,s)^T)_{k,l}$ $\forall k,l = 1,...,m$ and assume that:*

1. $\inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \frac{p(x)}{q_\theta(x)}\ddot{\ell}_\theta(x,s)_{k,l} > -\infty \ \forall k,l = 1,\dots,m.$
2. $\inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \left(\frac{p(x)}{q_\theta(x)}\right)^2 \dot{\ell}_\theta(x,s)\dot{\ell}_\theta(x,s)^T_{k,l} > -\infty \ \forall k,l = 1,\dots,m.$
3. $\iint L_{i,k,l}(x,s)P(dx,ds) < +\infty$, $\forall i \in \{1,2\}$, $\forall k,l = 1,...,m.$
4. $\sup_{\theta\in\Theta}\iint \frac{L_i(x,s)^2 p(x)}{q_\theta(x)}P(dx,ds) < +\infty, \ i \in \{1,2\}.$

Then the sequence $\widehat{G}_n^{-1}$ converges to $G_{\theta_*}^{-1}$ in probability.

By Theorem 2: $n(\tilde{\theta}_n - \theta_*)^T G_{\theta_*}^{-1}(\tilde{\theta}_n - \theta_*) \to \chi^2(m)$. Using Slutsky's lemma, we have the following convergence in distribution:

$$n(\tilde{\theta}_n - \theta_*)^T \widehat{G}_n^{-1}(\tilde{\theta}_n - \theta_*) \to \chi^2(m) \ . \tag{19}$$

Equation (19) allows to define asymptotic confidence ellipsoids that will be defined precisely in Section 3.3.

3.3. Stopping criterion using a statistical hypothesis test

The estimation of the generalization error without a validation set is often based on Cross Validation. When AIS is used, the data points (X_i, S_i) are no longer independent and identically distributed. We propose to use a stopping criterion that ensures that asymptotic normality is reached and then to use the asymptotic confidence ellipsoids determined above, which will then provide reliable confidence ellipsoids for the quantities of interest.

Theorem 4. *[Centered asymptotic normality of the loss gradient] Assume the same hypotheses as in Theorem 2 and:*

1. $\inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \frac{p(x)}{q_\theta(x)} \ddot{\ell}_{\theta_*}(x,s)_{k,l} > -\infty \ \forall k,l = 1,\dots,m.$

Consider two independent datasets $\mathcal{D}_1 = (X_{i,1}, S_{i,1})_{i\in\{1,\dots,n\}}$ *and* $\mathcal{D}_2 = (X_{i,2}, S_{i,2})_{i\in\{1,\dots,n\}}$ *generated with AIS strategy. Let* $\tilde{R}_{n,j}$ *be the weighted loss for* $\mathcal{D}_j$ *for* $j = 1, 2$ *defined as in (11). Denote:*

$$\tilde{\theta}_{n,j} = \operatorname*{argmin}_{\theta\in\Theta} \tilde{R}_{n,j}(\theta), \qquad j = 1, 2\,.$$

Then we have:

$$\sqrt{n}(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 8V(q_{\theta_*}, \dot{\ell}_{\theta_*}))$$

as $n \to +\infty$.

Denote

$$\widehat{W}_n = \frac{n}{8}(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1}))^T \widehat{V}_{n,12}^{-1}(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1})), \tag{20}$$

$$\widehat{V}_{n,12} = \frac{1}{2}\big(\widehat{V}_{n,1}(\tilde{\theta}_{n,1}) + \widehat{V}_{n,2}(\tilde{\theta}_{n,2})\big), \tag{21}$$

with $\widehat{V}_{n,j}$ the empirical estimator in equation (18) for the j-th AIS dataset $\mathcal{D}_j$ for $j = 1, 2$. By Theorem 4 and by Slutsky's lemma, $\widehat{W}_n$ converges weakly to $\chi^2(m)$. It is, therefore, possible to define a stopping criterion inspired by statistical test theory to check the asymptotic normality of $\tilde{\theta}_n$. Our stopping criterion is equivalent to the hypothesis test:

$$(\mathcal{H}_0) \ : \widehat{W}_n \text{ follows } \chi^2(m) \quad \text{against} \quad (\mathcal{H}_1) \ : \widehat{W}_n \text{ does not follow } \chi^2(m)\ . \tag{22}$$

For $\alpha \in (0, 1)$, we then consider the statistical test which rejects $(\mathcal{H}_0)$ if:

$$\widehat{W}_n > q^{1-\alpha}_{\chi^2(m)}\ , \tag{23}$$

where $q^{1-\alpha}_{\chi^2(m)}$ denotes the $(1-\alpha)$-quantile of the $\chi^2(m)$ distribution. Hence, this statistical test is of asymptotic level α. An apparent drawback of this stopping criterion is that it doubles the computational cost, due to the necessity of having two independent AIS estimators $\tilde{\theta}_{n,1}$ and $\tilde{\theta}_{n,2}$ to compute $\widehat{W}_n$. It is, however, possible to use the estimator

$$\tilde{\theta}_{n,12} = \frac{\tilde{\theta}_{n,1} + \tilde{\theta}_{n,2}}{2}, \tag{24}$$

which has an asymptotic variance that is half the one of $\tilde{\theta}_{n,1}$ and $\tilde{\theta}_{n,2}$ as shown by the following proposition. There is, therefore, no increase of the computational cost.

Theorem 5. *The sequence $\sqrt{n}(\tilde{\theta}_{n,12} - \theta_*)$ is asymptotically normal with mean zero and covariance matrix $\frac{1}{2}\ddot{r}(\theta_*)^{-1}V(q_{\theta_*}, \dot{\ell}_{\theta_*})(\ddot{r}(\theta_*)^{-1})^T$.*

Using Equation (19), it is possible to define the asymptotic confidence ellipsoids for the parameter θ_* at level α:

$$\hat{\mathcal{E}}_{n,\alpha} = \{\theta : (\theta - \tilde{\theta}_{n,12})^T \widehat{G}_{n,12}^{-1}(\theta - \tilde{\theta}_{n,12}) \leq q_{1-\alpha}^{\chi^2(m)}\} \ , \tag{25}$$

with

$$\widehat{G}_{n,12} = \frac{\widehat{\ddot{r}}_{n,12}^{-1}\widehat{V}_{n,12}(\widehat{\ddot{r}}_{n,12}^{-1})^T}{2} \ , \tag{26}$$

$$\widehat{\ddot{r}}_{n,12} = \frac{\widehat{\ddot{r}}_{n,1}(\tilde{\theta}_{n,1}) + \widehat{\ddot{r}}_{n,2}(\tilde{\theta}_{n,2})}{2} \ , \tag{27}$$

$\widehat{V}_{n,12}$ is defined by (21), and the estimators $\widehat{\ddot{r}}_{n,j}$ correspond to the estimator $\widehat{\ddot{r}}_n$ defined in Equation (17) with the j-th AIS sample $\mathcal{D}_j$. Once the stopping criterium $\widehat{W}_n < 1$ is fulfilled, we can claim that asymptotic normality is reached and the asymptotic confidence ellipsoid (25) can be used.

4. Practical implementation of AIS and comparison with a Random Sampling strategy

This section is primarily dedicated to explaining how to apply AIS to a real case study of seismic fragility curve estimation. For sake of clarity, we separate the implementation issue (section 4.1) from the initialization one (section 4.2). In addition, in order to demonstrate the superiority of the AIS strategy in comparison with a Random Sampling (RS) strategy, performance metrics are presented in section 4.3.

4.1. Practical implementation

This section is dedicated to explain how to apply AIS to a real case study of seismic fragility curve estimation. The classical parametric shape for seismic

fragility curve is $\mathcal{F} = \{\Phi(\frac{\log(\frac{IM}{\alpha})}{\beta}), (\alpha, \beta) \in (0, +\infty)^2\}$, with Φ the cumulative distribution function (cdf) of the standard normal distribution and $X = \log IM$ [1]. In the following we will use the notation $\theta = (\alpha,\ \beta)^T$. Remark that Θ is not compact. From an engineer perspective, it is possible to bound α and β from above and to bound α from below. However the neighborhood of 0 for β is problematic and thus, inspired by Bayesian inference theory [40], we consider a regularization $\Omega(\theta; \beta_{reg}) = \frac{\beta_{reg}}{\beta}$ to tackle this issue. The AIS loss in Algorithm 1 is replaced by:

$$\tilde{R}_{n,reg}(\theta; \beta_{reg}) = \frac{1}{n} \sum_{i=1}^{n} \frac{p(X_i)}{q_{\tilde{\theta}_{i-1}}(X_i)} (S_i - f_\theta(X_i))^2 + \frac{\Omega(\theta; \beta_{reg})}{n} \ . \tag{28}$$

This penalization parameter is motivated by the intrinsic difficulty of estimating the slope β of the lognormal model when β is small [41]. Fragility curves with small β are hard to distinguish due to the convergence towards a degenerate $0-1$ deterministic fragility curve. This phenomenon is magnified with a small-sized sample. Remarking that $\tilde{R}_{n,reg}(\theta; \beta_{reg}) \underset{n \to +\infty}{\sim} \tilde{R}_n(\theta)$, the convergence results for $\tilde{\theta}_n$ are guaranteed. The regularization only affects $\tilde{\theta}_n$ for small values of n. The regularization hyperparameter β_{reg} is determined using cross-validation. The computation of β_{reg} will be detailed in Section 4.2.

Moreover, using AIS with the optimal instrumental density q_θ directly could increase the variance if the density has light tails. We propose finally an Adaptive Defensive and Regularized Importance Sampling (ADRIS) strategy as illustrated in [42, 43]. The instrumental density becomes

$$\tilde{q}_{\theta,\varepsilon}(x) = \varepsilon p(x) + (1 - \varepsilon) q_\theta(x) \tag{29}$$

with $\varepsilon \in [0, 1]$. ε is a mixing parameter, between the original marginal pdf $p(x)$ and the instrumental one $q_\theta(x)$, meaning that one time out of $1/\varepsilon$ the element is drawn from the pdf $p(x)$. This distribution allows to bound the likelihood ratio:

$$\frac{p(x)}{\tilde{q}_{\theta,\varepsilon}(x)} = \frac{1}{\varepsilon + (1 - \varepsilon) \frac{q_\theta(x)}{p(x)}} < \frac{1}{\varepsilon} \ .$$

Thus the defensive strategy bounds the variance even if the likelihood ratio $\frac{p(x)}{q_\theta(x)}$ is large. In the numerical applications, we sample the seismic signals with

seismic intensity X from a very large dataset of precomputed seismic signals with the original distribution that is called the unlabeled pool. The number of seismic signals is considered very large (10^5 in our case) compared to the number of seismic signals used for fragility curve estimation, so that the theoretical results are still supposed valid. For ADRIS, the seismic signals are first sampled at random in the unlabeled pool, then an acceptance reject algorithm is used to sample with the instrumental density $\tilde{q}_{\theta,\varepsilon}$, the likelihood ratios are computed using Simpson's quadrature.

4.2. Initialization of ADRIS and RS strategies

4.2.1. Initialization of the ADRIS algorithm

A key aspect of ADRIS procedure is the initialization parameter $\tilde{\theta}_0$. As expected the closer $\tilde{\theta}_0$ is from the true parameter θ_* the faster ADRIS is in asymptotic normal regime. A naive approach is to get a small sample of size n_0 (e.g. $n_0 = 20$) $(X_i, S_i)_{i=1}^{n_0}$ from the original marginal density p of X and to compute the passive learning estimator $\widehat{\theta}_{n_0}$ (equation (4)). This crude estimation gives a sufficient initial parameter to start ADRIS.

A better approach is to consider a metamodel - in the broad sense - of the mechanical simulation. As often used by practitioners, a numerical resolution based on a modal base projection can be implemented to get an estimate of the fragility curve corresponding to the linear behavior of the structure of interest. It is then possible to get a huge amount of datapoints of the reduced model (e.g. an independent and identically distributed sample of $n_{\text{red}} = 10^3$–10^5 pairs $(X_i, S_{\text{red},i})_{i=1}^{n_{\text{red}}}$ where X_i is sampled with the original pdf p and $S_{\text{red},i}$ is the associated label obtained with the reduced model) in order to get a precise enough initial parameter $\tilde{\theta}_0$ approximated by $\widehat{\theta}_{n_{\text{red}}}$. Statistical metamodels could also be used such as Gaussian Processes [44] or Support Vector Machines [16].

In our applications reduced models are only used to give us prior knowledge on the fragility curve shape, encapsulated in the initial parameter of the ADRIS procedure. We then initialize ADRIS with a small sample of 20 datapoints with the instrumental density $\tilde{q}_{\tilde{\theta}_0,\varepsilon}$ (equation (29)). The regularization

parameter β_{reg} is then determined by minimizing the Leave One Out error on this initialization sample. We denote by $\tilde{\beta}_{reg}$ the corresponding regularization parameter.

4.2.2. Initialization of the RS algorithm

The initialization for the RS estimator is 20 datapoints sampled using the original distribution p, the penalization parameter is also computed using Leave One Out cross validation and the corresponding penalization parameter value is denoted by $\widehat{\beta}_{reg}$.

4.3. Performance metrics of the numerical benchmarks

The next section aims at assessing ADRIS performance in comparison with the RS strategy on test cases. For this, we here introduce performance metrics, inspired from [45, 46], to unveil the quality of our active learning procedure. In the following we denote respectively by $\widehat{\theta}_n$ and $\tilde{\theta}_n$ the RS and ADRIS estimators.

- the *Relative Standard Deviation*

$$\widetilde{\mathrm{RSD}}_n = \frac{\sqrt{\mathbb{V}[\tilde{R}_{n,reg}(\tilde{\theta}_n;\tilde{\beta}_{reg})]}}{\mathbb{E}[\tilde{R}_{n,reg}(\tilde{\theta}_n;\tilde{\beta}_{reg})]},$$

$$\widehat{\mathrm{RSD}}_n = \frac{\sqrt{\mathbb{V}[\widehat{R}_{n,reg}(\widehat{\theta}_n;\widehat{\beta}_{reg})]}}{\mathbb{E}[\widehat{R}_{n,reg}(\widehat{\theta}_n;\widehat{\beta}_{reg})]},$$

where $\widehat{R}_{n,reg}(\theta;\beta_{reg}) = \frac{1}{n}\sum_{i=1}^{n}(S_i - f_\theta(X_i))^2 + \frac{\Omega(\theta;\beta_{reg})}{n}$.

- the *Relative Biais*

$$\mathrm{RB}_n = \frac{|\mathbb{E}[\tilde{R}_{n,reg}(\tilde{\theta}_n;\tilde{\beta}_{reg})] - \mathbb{E}[\widehat{R}_n(\widehat{\theta}_n;\widehat{\beta}_{reg})]|}{\mathbb{E}[\tilde{R}_{n,reg}(\tilde{\theta}_n;\tilde{\beta}_{reg})]}.$$

- The *efficiency*

$$\nu_n = \frac{\mathbb{V}[\widehat{R}_{n,reg}(\widehat{\theta}_n;\widehat{\beta}_{reg})]}{\mathbb{V}[\tilde{R}_{n,reg}(\tilde{\theta}_n;\tilde{\beta}_{reg})]}.$$

A value of $\nu_n > 1$ shows that ADRIS has a smaller loss variance than RS. For simplicity we will only show the value of ν_n for the maximal sample size on each test case. The mean and variance above can be computed empirically

using replications of the two procedures (ADRIS and RS) when the test cases are simple and when it is possible to run many replicates of the ADRIS and RS samples.

These metrics are also defined with the empirical testing error:

$$Q(\theta;\beta_{reg}) = \frac{1}{n_t}\sum_{i=1}^{n_t}\left(S_i^{(t)} - f_\theta\left(X_i^{(t)}\right)\right)^2 + \frac{\Omega(\theta;\beta_{reg})}{n_t} \ ,$$

where $(X_i^{(t)}, S_i^{(t)})_{1\leq i\leq n_t}$ is a testing set (independent and identically distributed with the original distribution P). The performance metrics are thus also defined with the testing error by replacing $\widehat{R}_n$ or $\tilde{R}_n$ with Q and parameter $\widehat{\theta}_n$ or $\tilde{\theta}_n$.

5. Numerical results

To evaluate ADRIS efficiency, a numerical benchmark has been performed with three test cases with increasing complexity:

1) A synthetic test case with known fragility curve and probability distribution of the seismic log-intensity measure X,

2) a nonlinear elastoplastic oscillator with kinematic hardening subjected to synthetic signals generated from the modulated and filtered white-noise ground-motion model [26] as in [16],

3) an industrial test case of a nuclear facility's pipeline-system, submitted to the same artificial signals.

The oscillator test case aims to evaluate the effectiveness of the ADRIS strategy before its application to an industrial test case which is numerically much more costly. Moreover, since it well represents the essential features of the nonlinear responses of a large variety of real structures subjected to earthquakes, this test case allows to determine the value of the hyperparameter ε - thanks to a numerical benchmark - because there is no ad hoc procedure to do this.

5.1. A synthetic test case

Here we benchmark our methodology while having full knowledge of the true fragility curve. We generate 30,000 datapoints (X_i, S_i) with the fragility curve

$\mu(x) = \Phi(\frac{x-\log(\alpha_*)}{\beta_*})$ with $(\alpha_*, \beta_*) = (0.3, 0.4)$. The original marginal distribution of X is here a Gaussian distribution with mean $\log\left(\frac{\alpha_*}{3}\right)$ and variance β_*^2. The unlabeled pool consists of 20,000 datapoints X_i. 10,000 datapoints (X_i, S_i) will be our validation set for testing error estimation, using crude Monte Carlo. Figure 1 shows (i) the target fragility curve μ in dashed red line, (ii) a kernel density estimation of the density p based on the all dataset in green and (iii) a kernel density density estimation q of the 120 datapoints X_i obtained by ADRIS in red. Figure 2 plots the training error and testing error for 500 replications of the RS and ADRIS algorithms. The algorithms are initialized with 20 datapoints, 120 datapoints are extracted using ADRIS or RS from the unlabeled dataset.

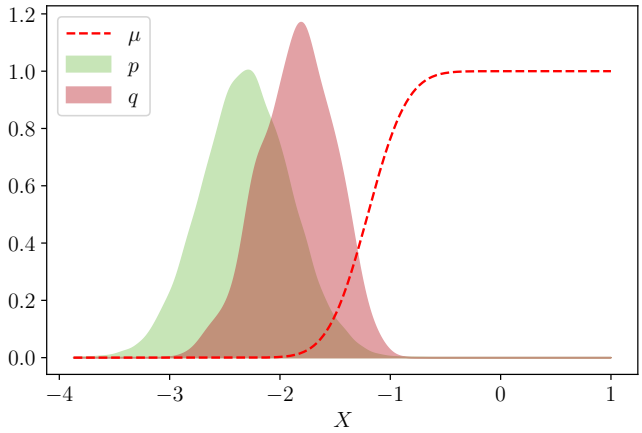


Figure 1: Synthetic test case with lognormal fragility curve with parameters $(\alpha_*, \beta_*) = (0.3, 0.4)$ and $X \sim \mathcal{N}(\frac{\alpha_*}{3}, \beta_*^2)$. Comparison of the original marginal density p of X with the empirical density q of the 120 datapoints X_i obtained by ADRIS for the synthetic test case.

As depicted by Figure 2 and Table 1, ADRIS does not seem to reduce the training error. This is normal because ADRIS selects seisms whose intensity measures maximize $\tilde{\ell}_\theta$, which can be seen as an marginalized training loss variance of the observation. In other words, as illustrated in Figure 1 with the density $q(x)$, ADRIS selects difficult points - typically values of x for which

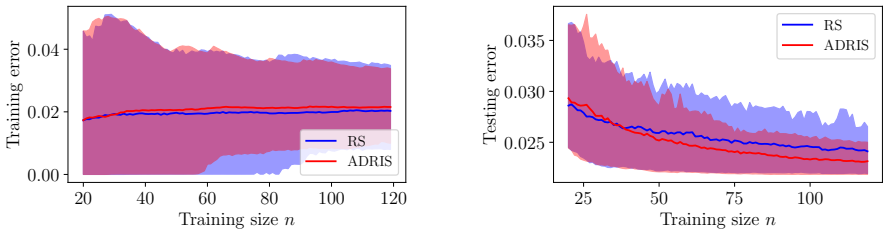


Figure 2: Numerical benchmark of the synthetic test case: The thick lines represent the mean loss for 500 replications, the shaded areas represents the range between the quantiles at 90% and 10%. For this case, the bias is known and is equal to $\mathbb{E}[\mu(X)(1-\mu(X))] \simeq 0.021$.

Table 1: Performance metrics for the synthetic test case for $n = 120$

	Train		Test	
	RS	ADRIS	RS	ADRIS
RSD_{120} (%)	53	46	13	8
ν_{120}	×	1.2	×	2.5
RB_{120} (%)	×	7	×	4

$\mu(x)$ takes values between 0 and 1 - and therefore the training error can be large because it is not representative of the generalization error. RS and ADRIS strategies really distinguish themselves on the testing error, which is smaller for ADRIS than for RS. Moreover, ADRIS quickly converges to the known bias equal to $\mathbb{E}[\mu(X)(1-\mu(X))] \simeq 0.021$. Finally the efficiency ν_{120} is 2.5, meaning that the testing error's confidence interval for ADRIS is 1.6 times smaller than for RS after 120 iterations.

5.2. A nonlinear oscillator

This test case aims to validate the overall strategy developed in this work, on a simple but representative case. The goal is to analyze empirically the performance of ADRIS for this test case, which will not be possible for complex

structures such as the one addressed in section 5.3. This section is therefore particularly comprehensive, starting from the initialization of the ADRIS algorithm up to the numerical verification of the theorems, while passing by the choice of ε.

275

5.2.1. Presentation of the oscillator

This second test case - illustrated in Figure 3 - is that of a single degree of freedom elasto-plastic oscillator which exhibits kinematic hardening. It was previously used as a simple mechanical model for fragility curve estimation in [20, 16]. Its equation of motion is:

$$\ddot{z}(t) + 2\beta\omega_L \dot{z}(t) + f_{NL}(t) = -s(t) \ ,$$

with $s(t)$ an artificial seismic signal, $\dot{z}(t)$ and $\ddot{z}(t)$ are respectively the velocity and acceleration of a unit mass of the oscillator, β is the damping ratio and ω_L the pulsation of the oscillator. Moreover, the nonlinear force f_{NL} is governed by
280 two parameters: the post-yield stiffness a and the yield displacement Y. We also define the maximal displacement $D = \max_{t\in[0,T]} z(t)$, where T is the duration of the seismic excitation. The failure state for this nonlinear oscillator is defined by the $\{0,1\}$-valued variable $S = \mathbb{1}_{(D>C)}$, where $C = 2Y$ is chosen such that $\mathbb{P}(D \geq C) \approx 10^{-1}$, which means that the underlying structure is "rather well
285 designed" with respect to the seismic scenario considered.

5.2.2. Initialisation of the ADRIS procedure

In this test case, for ADRIS initialization, we use the underlying elastic oscillator as a cheap model. The initialization parameter $\tilde{\theta}_0$ is approximated by
290 $\widehat{\theta}_{10^5}$ (equation (4)) using a 10^5-sized dataset. In addition, the PGA is first considered as IM. Even if the PGA is not known to be the best indicator, doing so helps to verify the relevance of the methodology in a "less favorable" case. Note that the influence of the IM on the results will be discussed in section 5.2.7. As shown in Figure 4, the parameter $\tilde{\theta}_0$ could be considered "close" to the true

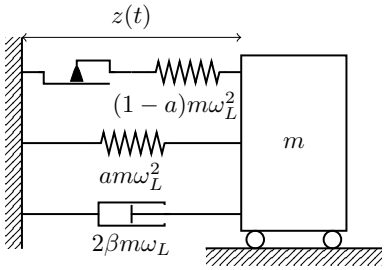


Figure 3: Elasto-plastic mechanical oscillator with kinematic hardening, with parameters $f_L = 5$ Hz and $\beta = 2\%$. The yield limit is $Y = 5.10^{-3}$ m and the post-yield stiffness is 20% of the elastic stiffness, hence $a = 0.2$.

parameter θ_*. Thus, 20 datapoints are queried on the nonlinear oscillator with the instrumental density $\tilde{q}_{\tilde{\theta}_0,\varepsilon}$ (equation (29)) before beginning launching the adaptive strategy.

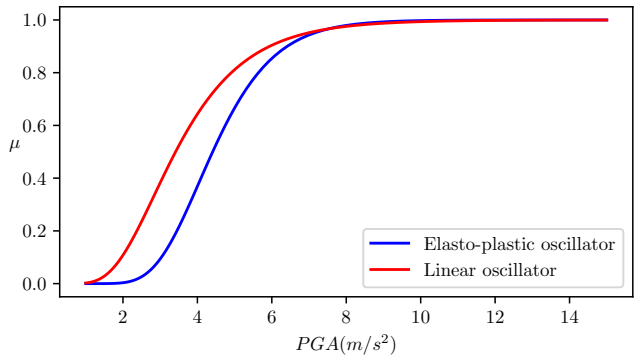


Figure 4: Fragility curves of the linear elastic oscillator and the nonlinear oscillator obtained by using least squares minimization on the total 10^5 synthetic seismic signals of the dataset.

5.2.3. Choice of ε

The choice of the defensive parameter value ε is cumbersome and there is no direct methodology for its choice. A benchmark on different values of ε is proposed in Table 2: remark that the test efficiency does not change between

10^{-2} and 10^{-3} however it is smaller when $\varepsilon = 10^{-1}$ meaning that the value $\varepsilon = 10^{-1}$ is too conservative. Accordingly, all the results will presented hereafter with a defensive parameter $\varepsilon = 10^{-3}$.

Table 2: Defensive parameter ε influence on the efficiency (IM = PGA).

ε	Train	Test
10^{-1}	1.3	1.2
10^{-2}	2.1	3.9
10^{-3}	2.2	3.3

5.2.4. Stopping criterion

Figure 5 shows the value of the test statistics $\widehat{W_n}$ for two independent ADRIS realizations, this motivates stopping the ADRIS algorithm at 100 sampled seismic signals, as the value of $\widehat{W}_n$ is inferior to the 95% quantile of the $\chi^2(2)$ distribution and we can then claim that we have reached asymptotic normality.

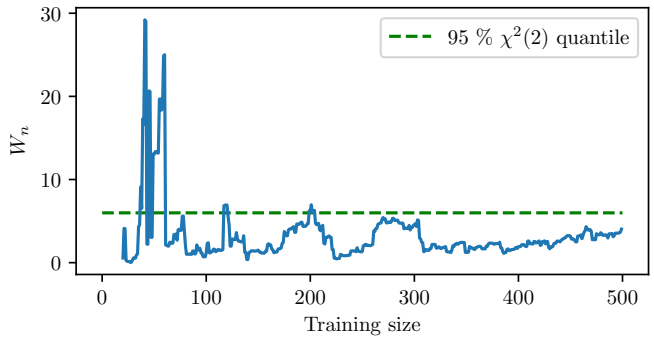


Figure 5: Values of the test statistic $\widehat{W}_n$ for two ADRIS realizations (IM = PGA).

5.2.5. Check of performances

In order to check the performances of the ADRIS algorithm, the unlabeled training set consists in 9.10^4 seismic signals and the testing set is composed of 10^4 signals. The benchmark study consists in 500 replications with 100 sampled seismic signals using ADRIS versus 120 for the RS strategy.

Figure 6 and Table 3 compare the ADRIS and RS training and testing errors. The mean training loss for the 500 replications is higher for ADRIS than for RS: indeed, the instrumental density is chosen to sample seismic signals that maximize the loss variance, resulting in a high training error. Remark that the ADRIS performance of the testing error is much better in terms of efficiency ν_{120} and RSD_{120}. The ADRIS confidence interval for testing error are thus 1.8 times smaller than for RS after 120 iterations. Moreover, the mean test error of ADRIS is also significantly smaller than for random sampling and quickly converges to the "minimal" error related to the term $\mathbb{E}[\mu(X)(1-\mu(X))]$ in (2).

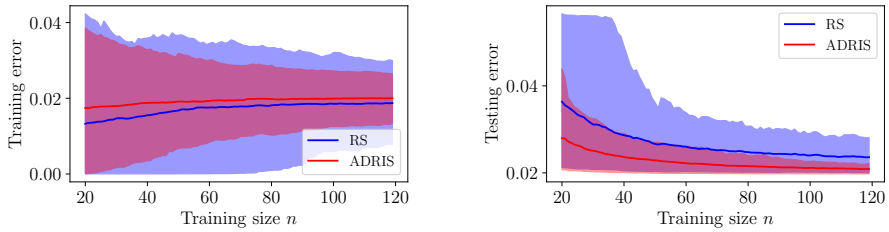


Figure 6: Results for the elasto-plastic mechanical oscillator with the PGA as IM (same notations as for the synthetic test case). ADRIS is initialized with 20 datapoints sampled from $q_{\tilde{\theta}_0,\varepsilon}$ with a defensive parameter $\varepsilon = 10^{-3}$. The confidence intervals are constructed from the empirical 90% and 10% quantiles of the training and testing losses for 500 replications.

The mean fragility curve and the 90% and 10% confidence interval are shown in Figure 7. The parametric fragility curve estimated with a dataset of 10^4 seismic ground motions and the nonparametric fragility curve estimated by k-means clustering on the PGA with a dataset of 10^5 seismic ground motions,

Table 3: Performance metrics for the elasto-plastic oscillator for $n = 120$ (IM = PGA).

	Train		Test	
	RS	ADRIS	RS	ADRIS
RSD_{120} (%)	47	30	19	11
ν_{120}	×	2.2	×	3.3
RB_{120} (%)	×	6	×	11

are also shown in order to validate both the model choice and the uncertainty reduction provided by ADRIS.

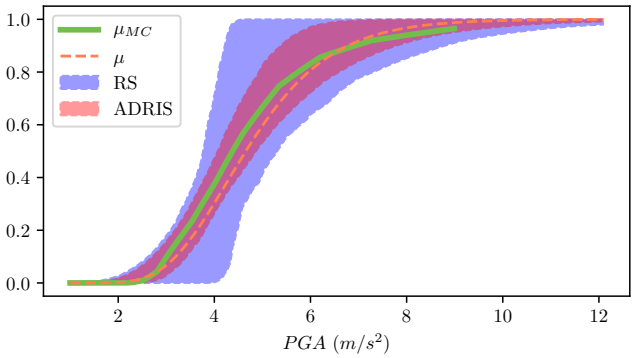


Figure 7: Estimation of the confidence intervals for random sampling and ADRIS for the nonlinear oscillator. Dashed orange curve and green curve are respectively the parametric estimation μ using 10^4 seismic ground motions and the k-means nonparametric estimation of the fragility curve using 10^5 seismic ground motions μ_{MC}. Red and blue areas correspond respectively to the 90% to 10% quantile range for the fragility curve dataset computed with ADRIS or RS. Remark that for the nonparametric fragility curve is only plotted for $PGA < 10m/s^2$ due to the lack of seismic signals with PGA above that threshold.

Figure 8 helps to visualize how ADRIS reduces the uncertainty of the fragility curve estimation: ADRIS is designed to sample seismic ground motions in the transition zone between 0 and 1 of the fragility curve, this phenomenon is responsible to the uncertainty reduction.

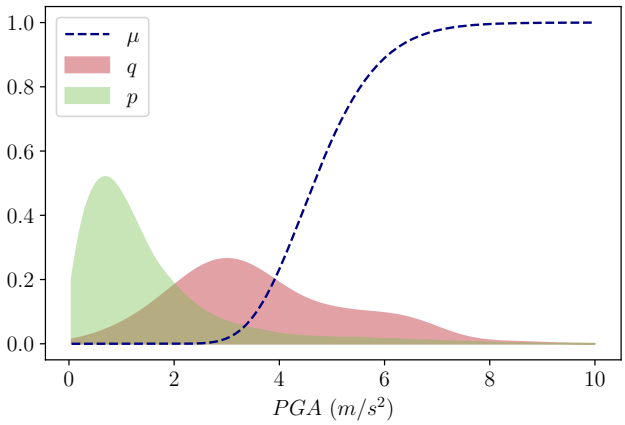


Figure 8: Comparison of the original marginal density p of PGA with the estimated density q of sampled PGA using ADRIS with $n = 100$ datapoints for the nonlinear oscillator.

5.2.6. Confidence ellipsoids

Numerical simulations have been carried out to check that the strategy proposed in the paper to build confidence ellipsoids is reliable. The experimental setup consists in 500 replicates with samples of size up to 1000. Each sample is obtained by ADRIS with a defensive parameter $\varepsilon = 10^{-3}$ and a penalization parameter chosen by cross validation. Two ADRIS samples are necessary to compute an asymptotic confidence ellipsoid with level $1 - \alpha = 0.9$ by Equation (25). The $B = 250$ replications of pairs of ADRIS are used to compute the Prediction Interval Coverage (PIC):

$$\widehat{\mathrm{PIC}}_n = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{\theta_* \in \hat{\mathcal{E}}_{n,\alpha}^{(b)}} \ ,$$

where $\hat{\mathcal{E}}_{n,\alpha}^{(b)}$ is the asymptotic confidence ellipsoid using the b-th replication of two independent runs of ADRIS procedure with size n. Because θ_* is unknown, we approximate it by $\widehat{\theta}_N$ for a very large $N = 10^5$. Figure 9 shows that for a training size n between 50 and 500, the PIC values are close to its theoretical and nominal value of 90%. This result validates the theoretical results of Section 3.2.

We emphasize that the stopping criterion $\widehat{W}_n$, illustrated in Figure 5, gives us at which sample size the ADRIS reaches asymptotic normality and thus at which sample size asymptotic confidence ellipsoids can be used.

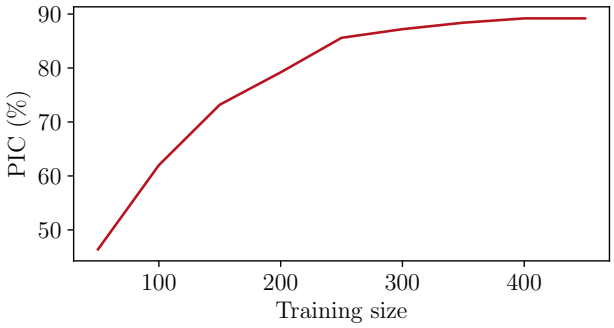


Figure 9: PIC value (for the confidence ellipsoid with level $1 - \alpha = 0.9$ for θ) as a function of the training size n of ADRIS. $B = 250$ ADRIS replications are used to estimate the PIC for each training size n.

5.2.7. Influence of the seismic intensity measure

The choice of the seismic IM is crucial for the accuracy of fragility curves estimates, especially when parametric models are concerned. With the PGA, a bias between the lognormal fragility curve and the k-means nonparametric fragility curve can indeed be seen in Figure 7. This phenomenon could be explained by the small correlation between maximal displacement of the oscillator during the seismic excitation and the PGA, which conveys small information about the seismic ground motion [32].

Confidence intervals for RS and ADRIS are thus computed again but using the spectral acceleration (SA) at 5 Hz and 2% damping ratio as the seismic IM. The results presented in Figure 10 show a reduction of the bias between the nonparametric and the parametric fragility curve. This illustrates that, for the class of structures and for the seismic signal generator considered in this study,

the parametric lognormal model has a better fit with the reference SA-based fragility curve than with reference PGA-based fragility curve.

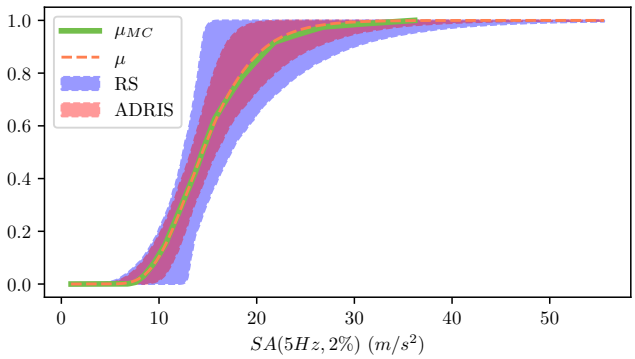


Figure 10: Estimation of the confidence intervals for RS and ADRIS for the nonlinear oscillator. Dashed orange curve and green curve are respectively the parametric estimation μ using 10^4 seismic ground motions and the k-means nonparametric estimation of the fragility curve using 10^5 seismic ground motions μ_{MC}. Red and blue areas correspond respectively to the 90% to 10% quantile range for the fragility curve dataset computed with ADRIS or RS. The seismic IM in this case is the spectral acceleration at 5 Hz and 2% damping ratio. Remark that the bias between the nonparametric fragility curve μ_{MC} and the parametric fragility curve μ is smaller than with the PGA for intensity measure.

5.2.8. Synthesis

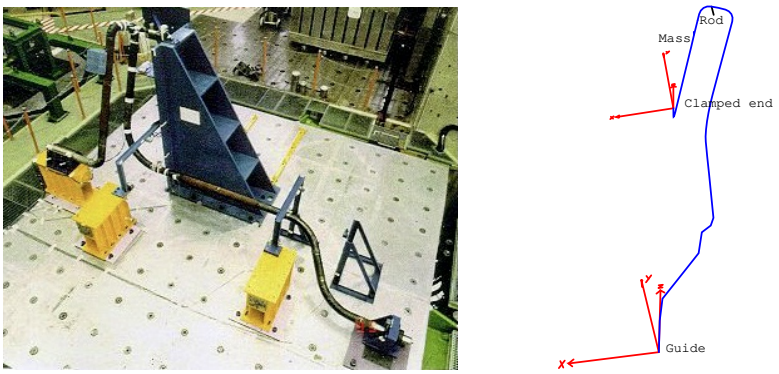
In this section, we have shown that the ADRIS-based methodology is (i) efficient to reduce the variance of the fragility curve estimation and (ii) can be applied regardless of the IM of interest. However, in practice, it is more suitable to use an IM as correlated as possible to the response of the structure to minimize potential biases due to the use of a parametric model. In addition, we have shown that, if the computation times allow it, it is possible to know when to stop the ADRIS algorithm, in order to build an asymptotic confidence ellipsoids.

The piping systems (or some of their sections) are part of the class of structures for which the global nonlinear seismic behavior can be well represented

by nonlinear oscillators. So, we consider for the following test case that, after 100 iterations, for $\varepsilon = 10^{-3}$, the asymptotic normality will be reached regarding the use of asymptotic confidence ellipsoids. Initialization will follow the same procedure as for the nonlinear oscillator. In addition, SA will be chosen as IM for the fragility curve.

5.3. Industrial test case: safety water supply pipe in a water pressured nuclear reactor

The following test case corresponds to a piping system which is a simplified part of a secondary line of a French Pressurized Water Reactor. The numerical model was validated based on seismic tests performed on the shaking table Azalee of the EMSI laboratory of CEA/Saclay. The experimental program, called ASG program, and the main results are outlined in Ref. [47]. In Figure 11a a view of the mock-up mounted on the shaking table is shown. The Finite Element (FE) model, based on beam elements, is depicted in Figure 11b.



(a) (b)

Figure 11: (a) Overview of the ASG mock-up on the shaking table and (b) ASG FE model

The mock-up is a 114.3 mm outside diameter and 8.56 mm thickness pipe with a 0.47 elbow characteristic parameter, in carbon steel TU42C, filled with water without pressure. It contains three elbows and a mass modeling a valve (120 kg) which corresponds to more than 30% of the specimen total mass. As

shown in Figure 11b, one end of the mock-up is clamped whereas the other is supported by a guide in order to prevent the displacements in the X and Y directions. Additionally, a rod is placed on the top of the specimen in order to limit the mass displacements in the Z direction. In the tests, excitation act in the X direction.

Numerical comparisons are carried out with the homemade FE code CAST3M [48]. Concerning the FE model, the boundary conditions are adjusted in order to obtain shapes and frequencies similar of those of the first two eigenmodes of the mock-up in the X and Y directions, respectively at 5.1 Hz and 6.6 Hz. As measured in the experiments, a critical damping ratio of 1% is considered for these two eigenmodes with a damping Rayleigh assumption. Finally, regarding the nonlinear constitutive law of the material, a bilinear law exhibiting kinematic hardening is used to reproduce the overall non-linear behavior of the mock-up with satisfactory agreement compared to the results of seismic tests [47].

In the context of this test case, the yield stress of the bilinear law is equal to 3 10^8 Pa, the Young modulus is equal to 1.92 10^{11} Pa whereas the hardening modulus is equal to 4.3 10^8 Pa. Moreover, since for the synthetic signals considered in this work (the same as those used in the reference [16] and in the second test case of this paper) the piping system remains in the linear domain, they are filtered by a fictitious linear single-mode building at 5 Hz and damped at 2%. Finally, we consider excessive out-of-plane rotation of the elbow located near the clamped end of the mock-up as failure criterion, as recommended in [49]. Since the weight of the mass is not completely taken up by the mechanical assembly, the overall behavior of the mock-up exhibits ratcheting.

In the following, the random variable R corresponds to the maximum of the out-of-plane rotation of the elbow. The binary variable which indicates the failure state is defined by $S = \mathbf{1}_{R>C}$ where C is the admissible rotation in degree. In our case, $C = 4.38°$. This value of admissible rotation is the 90%-level quantile from a sample of 1000 mechanical simulations, it is consistent with an industrial case in which the failure is a "rare event". The ADRIS procedure is initialized by considering the linear FE model and a numerical resolution

based on a modal base projection. Figure 12 and Table 4 are the results of a numerical benchmark consisting of 50 replications of 100 signals sampled using ADRIS with a defensive parameter $\varepsilon = 10^{-3}$ and 120 signals chosen at random in the unlabeled pool. The numbers of chosen signals have the same order of magnitude as for the previous test case of the nonlinear oscillator. Figure 13 illustrates the uncertainty reduction provided by ADRIS on the fragility curve estimate. Motivated by the detailed results obtained for the nonlinear oscillator and presented in the previous section, the fragility curve of the piping system is here expressed as a function of the pseudo-spectral acceleration of the initial set of the synthetic signals (i.e not filtered signals), calculated at 5 Hz and 1% damping ratio. Note that Figure 13 suggests that for the RS strategy, even with $n = 120$ points, it is possible to obtain samples for which a β estimate is equal to 0 (the 10% quantile is vertical around SA $\simeq 11m/s^2$), which ADRIS avoids.

Table 4: Performance metrics for the ASG piping system for $n = 120$

	Train		Test	
	RS	ADRIS	RS	ADRIS
RSD_{120} (%)	40.5	40	24.1	11.6
ν_{120}	×	1.2	×	5.1
RB_{120} (%)	×	9.5	×	20

6. Conclusion

In this paper, we have introduced an original methodology to improve estimation accuracy of fragility curve without increasing the sample size, thanks to an active learning strategy of adaptive importance sampling. Defensive strategy has been implemented to control the likelihood ratio and the possible increase of the training loss variance in the early steps. We use a penalized least square loss to prevent an identifiability issue for the slope parameter of the fragility curve. We define a stopping criterium that indicates asymptotic normality of

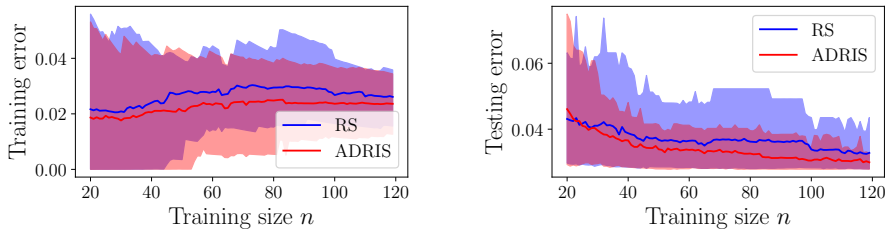


Figure 12: Numerical benchmark of the ASG piping system. The confidence interval of the train and test losses are computed with empirical quantiles of levels 10% and 90% on the 50 replications.

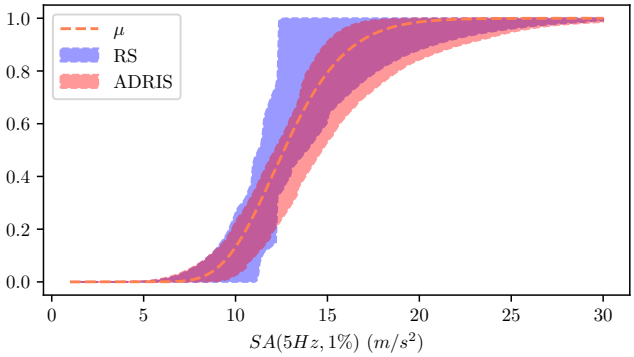


Figure 13: Estimation of the confidence intervals for RS and ADRIS for the ASG piping system. Red and blue areas correspond respectively to the range between the 10% and 90% quantiles of the fragility curve dataset generated with ADRIS or random sampling with $n = 100$ training datapoints. The dashed orange curve corresponds to a fragility curve estimation using least squares minimization with a dataset of 1000 seismic ground motions and FE simulations of the piping system.

the estimator and provide asymptotic confidence intervals and ellipsoids. We illustrate our active learning procedure in numerical examples, from a synthetic case to a FE mechanical simulation of a piping system of a French Pressurized Water Reactor.

Acknowledgements

This research was supported by CEA (French Alternative Energies and Atomic Energy Commission) and SEISM Institute (www.institut-seism.fr/en/).

Appendix A. Lemma

Lemma 2 (A law of large numbers)**.** *Let $U_n, n \geq 1$ be a sequence of almost surely positive random variables. Let $S_n = \sum_{i=1}^{n} U_i$ be such that:*

$$\frac{\mathbb{E}[S_n]}{n} \overset{n\to+\infty}{\longrightarrow} l, \qquad Var(S_n) \leq cn, \qquad \forall n \geq 1,$$

for some real numbers $c \geq 0$, $l \geq 0$. We then have

$$\frac{S_n}{n} \to l \ a.s.$$

Proof. See Theorem in Appendix B.3 in [22]. □

Theorem 6 (Uniform law of large numbers)**.** *Let Θ be a compact set of $\mathbb{R}^m$. Consider a sequence of random functions $\theta \mapsto H_i(\theta)$, $\forall i \geq 1$, such that:*

- ***H1*** *There exists a stochastic process $H(\theta)$, such that for all $\theta \in \Theta$*

$$\frac{1}{n}\sum_{i=1}^{n} H_i(\theta) \to \mathbb{E}[H(\theta)] \ a.s.$$

In addition

$$\mathbb{E}\left[\sup_{\theta\in\Theta} |H(\theta)|\right] < +\infty,$$

and for any ball $\mathcal{B}$ of center θ_0

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{\theta\in\mathcal{B}} |H_i(\theta) - H_i(\theta_0)| \to \mathbb{E}\left[\sup_{\theta\in\mathcal{B}} |H(\theta) - H(\theta_0)|\right].$$

- ***H2*** *The function $\theta \to H(\theta)$ is continuous.*

Then the function $h(\theta) = \mathbb{E}[H(\theta)]$ is continuous and

$$\sup_{\theta\in\Theta} \left| h(\theta) - \frac{1}{n}\sum_{i=1}^{n} H_i(\theta) \right| \xrightarrow{n\to+\infty} 0 \ a.s.$$

Proof. See Theorem 5 in [22]. □

Appendix B. Proof of Theorem 1

We follow the lines of the proof of Theorem 2 in [22]. Using Theorem 5.7 in [50], we just need to check the two following conditions:

$$\sup_{\theta\in\Theta} |R_n(\theta) - r(\theta)| \overset{n\to+\infty}{\to} 0 \; a.s. \,, \tag{B.1}$$

$$\forall \varepsilon > 0 \,, \inf_{\theta\in\Theta, \|\theta-\theta_*\|>\varepsilon} \iint \ell_\theta(x,s) P(dx,ds) > \iint \ell_{\theta_*}(x,s) P(dx,ds) \,. \tag{B.2}$$

Since Θ is compact, condition (B.2) is satisfied due to the integrability of L which implies by the Lebesgue theorem that $\theta \to \iint \ell_\theta(x,s)P(dx,ds)$ is continuous. For the condition (B.1), we shall prove that $\tilde{R}_n(\theta)$ satisfies a uniform law of large numbers. We will show that Theorem 6 is satisfied. First of all, we use Lemma 2 in Appendix A to show that $\tilde{R}_n(\theta)$ satisfies a law of large numbers for all $\theta \in \Theta$. Using the same notations, we define $S_n(\theta) = \sum_{i=1}^{n} H_i(\theta)$ with $H_i(\theta) = \frac{p(X_i)\ell_\theta(X_i,S_i)}{q_{\tilde{\theta}_{i-1}}(X_i)}$ and $H(\theta) = \iint \ell_\theta(x,s)P(dx,ds)$. Denote by $\mathcal{F}_i$ the sigma algebra generated by $(X_j,S_j)_{j=1}^{i}$. Since $\mathbb{E}[H_i(\theta)|\mathcal{F}_{i-1}] = \int \ell_\theta(x,s)P(dx,ds)$, we get

$$\mathbb{E}[S_n(\theta)] = n \iint \ell_\theta(x,s)P(dx,ds) \;,$$

and

$$Var(S_n) = \sum_{i=1}^{n} Var(H_i(\theta)) \leq \sum_{i=1}^{n} \mathbb{E}[H_i(\theta)^2] \,,$$

the previous equality is derived from $Cov(H_j(\theta), H_i(\theta)) = 0$ for $j < i$. But for each i,

$$\begin{aligned}\mathbb{E}[H_i(\theta)^2] = \mathbb{E}\left[\frac{p(X_i)^2\ell_\theta(X_i,S_i)^2}{q_{i-1}(X_i)^2}\right] &= \iint \frac{p(x)\ell_\theta(x,s)^2}{q_{\tilde{\theta}_{i-1}}(x)} P(dx,ds) \\ &\leq \sup_{\theta'\in\Theta} \iint \frac{p(x)\ell_\theta(x,s)^2}{q_{\theta'}(x)} P(dx,ds) \;.\end{aligned}$$

This proves the law or large numbers for $\tilde{R}_n(\theta)$ for all $\theta \in \Theta$. The next step is to prove the uniform law of large numbers. For this, consider $\mathcal{B}$ a ball of center $\theta_0 \in \Theta$ and the random series $S_n = \sum_{i=1}^{n} U_i$ with $U_i = \sup_{\theta\in\mathcal{B}} |H_i(\theta) - H_i(\theta_0)|$. Thus

$$\mathbb{E}[S_n] = n \iint \sup_{\theta\in\mathcal{B}} |\ell_\theta(x,s) - \ell_{\theta_0}(x,s)| P(dx,ds) \;,$$

$$Var(S_n) = \sum_{i=1}^{n} Var(U_i) \leq \sum_{i=1}^{n} \mathbb{E}[U_i^2] \ ,$$

and

$$\begin{aligned}\mathbb{E}[U_i^2] &= \iint \frac{p(x) \sup_{\theta \in \mathcal{B}} |\ell_\theta(x,s) - \ell_{\theta_0}(x,s)|^2}{q_{\tilde{\theta}_{i-1}}(x)} P(dx,ds) \\ &\leq 2 \sup_{\theta \in \Theta} \iint \frac{p(x) \sup_{\theta \in \mathcal{B}} \ell_\theta(x,s)^2}{q_\theta(x)} P(dx,ds).\end{aligned}$$

Hence we have

$$\frac{1}{n} \sum_{i=1}^{n} H_i(\theta) \to \mathbb{E}[\ell_\theta(X,S)] \, .$$

The hypothesis **(H2)** of Theorem 6 is verified. As a consequence, $\tilde{\theta}_n$ is consistent.

Appendix C. Proof of Theorem 2

The asymptotic normality of an estimator built such as $\tilde{\theta}_n$ is based on the following arguments highlighted in Theorem 5.41 of [50]:

- **(H1)** The random function $\sqrt{n}\Psi_n(\theta_*) = \sqrt{n}(\dot{\tilde{R}}_n(\theta_*) - \dot{r}(\theta_*))$ converges in law to a centered Gaussian distribution with covariance V_{θ_*}.
- **(H2)** The random function $\dot{\Psi}_n(\theta_*)$ converges in probability to its expectation $\mathbb{E}[\dot{\Psi}_n(\theta_*)] = \mathbb{E}[\ddot{\ell}_{\theta_*}(X,S)]$
- **(H3)** The random function $\ddot{\Psi}_n(\theta_n)$ is bounded in probability for θ_n a deterministic sequence in a neighborhood of θ_*.

Of course, we need all the quantities above to be properly defined, hence we have to restrict ourselves to a loss function $\theta \mapsto \ell_\theta$ that is smooth enough, such as the quadratic loss. We will use the property that $\sqrt{n}\Psi_n(\theta_*)$ is a martingale, using Corollary 3.1 of [51], with $\Delta_i = \frac{p(X_i)\dot{\ell}_{\theta_*}(X_i,S_i)}{q_{\tilde{\theta}_{i-1}}(X_i)} - \dot{r}(\theta_*)$ the martingale increments and $U_{n,i} = \frac{1}{\sqrt{n}}\Delta_i$. Denote by $\mathcal{F}_n$ the σ-algebra generated by the family $(X_j, S_j)_{1 \leq j \leq n}$, the stochastic process $U_i, i \geq 1$ must satisfy

$$\sum_{i=1}^{n} \mathbb{E}[U_{n,i} U_{n,i}^T | \mathcal{F}_{i-1}] \to V_{\theta_*} \text{ in probability} \ , \tag{C.1}$$

$$\forall \varepsilon > 0,\ \sum_{i=1}^{n} \mathbb{E}[U_{n,i} U_{n,i}^T \mathbf{1}_{||U_{n,i}|| > \varepsilon} | \mathcal{F}_{i-1}] \to 0 \text{ in probability .} \qquad \text{(C.2)}$$

For (C.1), notice that

$$n\mathbb{E}[U_{n,i} U_{n,i}^T | \mathcal{F}_{i-1}] = \mathbb{E}\left[\frac{p(X) \dot{\ell}_{\theta_*}(X,S) \dot{\ell}_{\theta_*}(X,S)^T}{q_{\tilde{\theta}_{i-1}}(X)} | \tilde{\theta}_{i-1} \right],$$

where the expectation is taken with respect to (X, S) following the P distribution. Therefore, by Cesaro lemma, the convergence to V_{θ_*} is ensured. For the Lindberg condition (C.2), notice that

$$\sum_{i=1}^{n} \mathbb{E}[U_{n,i} U_{n,i}^T \mathbf{1}_{||U_{n,i}|| > \varepsilon} | \mathcal{F}_{i-1}] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\Delta_i \Delta_i^T \mathbf{1}_{||\Delta_i|| > \varepsilon \sqrt{n}} | \mathcal{F}_{i-1}] ,$$

and that

$$\mathbb{E}[\Delta_i \Delta_i^T \mathbf{1}_{||\Delta_i|| > \varepsilon \sqrt{n}} | \mathcal{F}_{i-1}] \leq (\varepsilon \sqrt{n})^2 \mathbb{E}[\mathbf{1}_{||\Delta_i|| > \varepsilon \sqrt{n}} | \mathcal{F}_{i-1}] ,$$

by applying the Markov inequality to the order $2 + \eta$ on the r.h.s. and the triangular inequality

$$\mathbb{E}[\Delta_i \Delta_i^T \mathbf{1}_{||\Delta_i|| > \varepsilon \sqrt{n}} | \mathcal{F}_{i-1}] \leq \frac{\sup_{i \in \mathbb{N}} \mathbb{E}[|| \frac{p(X) \dot{\ell}_{\theta_*}(X,S)}{q_{\tilde{\theta}_{i-1}}(X)} ||^{2+\eta} | \mathcal{F}_{i-1}] + \|\dot{r}(\theta)\|^{2+\eta}}{(\varepsilon \sqrt{n})^{\eta}} .$$

Hence, $\sqrt{n} \Psi_n(\theta_*)$ converges to a centered Gaussian distribution with covariance matrix equal to V_{θ_*}. Theorem 2.18 in [51] ensures that **(H2)** and **(H3)** are verified by the assumptions 2), 3) and 4) so that $\dot{\Psi}_n(\theta_*)$ converges toward the matrix $\ddot{r}(\theta_*)$. The sequence $\sqrt{n}(\tilde{\theta}_n - \theta_*)$ is asymptotically normal with mean zero and covariance matrix $\ddot{r}(\theta_*)^{-1} V_{\theta_*} (\ddot{r}(\theta_*)^{-1})^T$.

Appendix D. Proof of Theorem 3

The main arguments are the convergence in probability $\sqrt{n}(\tilde{\theta}_n - \theta_*) \to 0$ and the Taylor expansion around $\tilde{\theta}_n$ (using $\dot{\tilde{R}}_n(\tilde{\theta}_n) = 0$):

$$\begin{aligned} \sqrt{n}(\tilde{R}_n(\tilde{\theta}_n) - r(\theta_*)) =& \sqrt{n}(\tilde{R}_n(\theta_*) - r(\theta_*)) \\ &+ \sqrt{n} \int_0^1 (t-1)(\tilde{\theta}_n - \theta_*)^T \ddot{\tilde{R}}_n(\tilde{\theta}_n + t(\theta_* - \tilde{\theta}_n))(\tilde{\theta}_n - \theta_*) dt. \end{aligned}$$

Thanks of the boundedness condition of Theorem 2, under the hypothesis of Theorem 2, for all $t \in [0,1]$, $\ddot{\tilde{R}}_n(\theta_* + t(\tilde{\theta}_n - \theta_*))$ is bounded in probability, thus thanks to Corollary 3.1 of [51] and Slutsky's lemma, $\sqrt{n}(\tilde{R}_n(\theta_*) - r(\theta_*))$ is asymptotically normal.

Appendix E. Proof of Lemma 1

The result comes from the uniform convergence of $\widehat{G}_n(\theta) = \widehat{\ddot{r}}_n(\theta)^{-1}\widehat{V}_n(\theta)(\widehat{\ddot{r}}_n(\theta)^{-1})^T$ to G_θ for θ in a neighborhood of θ_*. It boils down to prove uniform convergence of $\widehat{\ddot{r}}_n(\theta)$ and $\widehat{V}_n(\theta)$. The proof is in the same spirit as the proof of Theorem 1. We proceed coordinate by coordinate defining $H_i(\theta)_{k,l} = \frac{p(X_i)}{q_\theta(X_i)}\ddot{\ell}_\theta(X_i,S_i)_{k,l} - \inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \frac{p(x)}{q_\theta(x)}\ddot{\ell}_\theta(x,s)_{k,l}$ to prove uniform convergence of $\widehat{\ddot{r}}_n(\theta)$ and $H_i(\theta)_{k,l} = \left(\frac{p(X_i)}{q_\theta(X_i)}\right)^2 \dot{\ell}_\theta(X_i,S_i)\dot{\ell}_\theta(X_i,S_i)^T_{k,l} - \inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \left(\frac{p(x)}{q_\theta(x)}\right)^2 \dot{\ell}_\theta(x,s)\dot{\ell}_\theta(x,s)^T_{k,l}$ for $\widehat{V}_n(\theta)$. Assumptions 3) and 4) ensure the uniform convergence using the proof technique of Theorem 1.

Appendix F. Proof of Theorem 4

The proof relies on the Taylor expansions of $\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2})$ and $\dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1})$ around the parameter value θ_*:

$$\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) = \dot{\tilde{R}}_{n,1}(\theta_*) + \ddot{\tilde{R}}_{n,1}(\theta_*)(\tilde{\theta}_{n,2} - \theta_*) + o\left(\|\tilde{\theta}_{n,2} - \theta_*\|\right).$$

$$\dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1}) = \dot{\tilde{R}}_{n,2}(\theta_*) + \ddot{\tilde{R}}_{n,2}(\theta_*)(\tilde{\theta}_{n,1} - \theta_*) + o\left(\|\tilde{\theta}_{n,1} - \theta_*\|\right).$$

Using hypothesis 1) of Theorem 4 and hypothesis 3) of Theorem 1, we can apply lemma 2 to prove convergence of $\ddot{\tilde{R}}_{n,1}(\theta_*)$ and $\ddot{\tilde{R}}_{n,2}(\theta_*)$ to $\ddot{r}(\theta_*)$ in the same spirit as for the proof of Theorem 1. We proceed coordinate by coordinate, defining $H_i(\theta_*)_{k,l} = \frac{p(X_i)}{q_{\tilde{\theta}_{i-1}(X_i)}}\ddot{\ell}_{\theta_*}(X_i,S_i)_{k,l} - \inf_{(\theta,x,s)\in\Theta\times\mathcal{X}\times\{0,1\}} \frac{p(x)}{q_\theta(x)}\ddot{\ell}_{\theta_*}(x,s)_{k,l}$. Remark that $H_i(\theta_*)_{k,l} \geq 0$, hence we can apply Lemma 2 to obtain the desired convergence. Moreover, the Taylor expansions of $\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,1})$ and $\dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,2})$ write:

$$0 = \dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,1}) = \dot{\tilde{R}}_{n,1}(\theta_*) + \ddot{r}(\theta_*)(\tilde{\theta}_{n,1} - \theta_*) + o\left(\|\tilde{\theta}_{n,1} - \theta_*\|\right),$$

$$0 = \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,2}) = \dot{\tilde{R}}_{n,2}(\theta_*) + \ddot{r}(\theta_*)(\tilde{\theta}_{n,2} - \theta_*) + o\left(\|\tilde{\theta}_{n,2} - \theta_*\|\right).$$

Finally, the Taylor expansion of $\sqrt{n}(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1}))$ writes:

$$\begin{aligned}
&\sqrt{n}\big(\dot{\tilde{R}}_{n,1}(\tilde{\theta}_{n,2}) - \dot{\tilde{R}}_{n,2}(\tilde{\theta}_{n,1})\big) \\
&= \sqrt{n}\big(\dot{\tilde{R}}_{n,1}(\theta_*) - \dot{\tilde{R}}_{n,2}(\theta_*) + \ddot{r}(\theta_*)(\tilde{\theta}_{n,2} - \theta_*) - \ddot{r}(\theta_*)(\tilde{\theta}_{n,1} - \theta_*)\big) + o_P(1) \\
&= 2\sqrt{n}(\ddot{\tilde{R}}_{n,1}(\theta_*) - \ddot{\tilde{R}}_{n,2}(\theta_*)) + o_P(1), \qquad \text{(F.1)}
\end{aligned}$$

because $\sqrt{n}\|\tilde{\theta}_{n,j} - \theta_*\| = O_P(1)$ for $j = 1, 2$. The right-hand side of Eq. (F.1) weakly converges towards the centered Gaussian distribution with covariance matrix $8V(q_{\theta_*}, \dot{\ell}_{\theta_*})$.

References

[1] R. Kennedy, C. Cornell, R. Campbell, S. Kaplan, H. Perla, Probabilistic seismic safety study of an existing nuclear power plant, Nuclear Engineering and Design 59 (2) (1980) 315 – 338.

[2] R. Kennedy, M. Ravindra, Seismic fragilities for nuclear power plant risk studies, Nuclear Engineering and Design 79 (1) (1984) 47 – 68.

[3] A. Ghobarah, Performance-based design in earthquake engineering: state of development, Engineering Structures 23 (8) (2001) 878 – 884.

[4] J. E. Richardson, G. Bagchi, R. J. Brazee, The seismic safety margins research program of the u.s. nuclear regulatory commission, Nuclear Engineering and Design 59 (1) (1980) 15 – 25.

[5] L. Eads, E. Miranda, H. Krawinkler, D. G. Lignos, An efficient method for estimating the collapse risk of structures in seismic regions, Earthquake Engineering & Structural Dynamics 42 (1) (2013) 25–41.

[6] J. E. Padgett, R. DesRoches, Sensitivity of seismic response and fragility to parameter uncertainty, Journal of Structural Engineering 133 (12) (2007) 1710–1718.

[7] A. Quilligan, A. O'Connor, V. Pakrashi, Fragility analysis of steel and concrete wind turbine towers, Engineering Structures 36 (2012) 270–282.

[8] B. R. Ellingwood, K. Kinali, Quantifying and communicating uncertainty in seismic risk assessment, Structural Safety 31 (2) (2009) 179 – 187.

[9] A. D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural Safety 31 (2) (2009) 105 – 112.

[10] R. Y. Rubinstein, D. P. Kroese, Simulation and the Monte Carlo Method, 2nd Edition, Wiley Publishing, 2008.

[11] M. Shinozuka, M. Q. Feng, J. Lee, T. Naganuma, Statistical analysis of fragility curves, Journal of Engineering Mechanics 126 (12) (2000) 1224–1231.

[12] B. R. Ellingwood, Earthquake risk assessment of building structures, Reliability Engineering & System Safety 74 (3) (2001) 251 – 262.

[13] T. K. Mandal, S. Ghosh, N. N. Pujari, Seismic fragility analysis of a typical indian PHWR containment: Comparison of fragility models, Structural Safety 58 (2016) 11–19. doi:https://doi.org/10.1016/j.strusafe.2015.08.003.
URL http://www.sciencedirect.com/science/article/pii/S0167473015000636

[14] F. Wang, C. Feau, Influence of Input Motion's Control Point Location in Nonlinear SSI Analysis of Equipment Seismic Fragilities: Case Study on the Kashiwazaki-Kariwa NPP, Pure and Applied Geophysicsdoi:https://doi.org/10.1016/j.engstruct.2018.02.024.

[15] I. Gidaris, A. A. Taflanidis, G. P. Mavroeidis, Kriging metamodeling in seismic risk assessment based on stochastic ground motion models, Earthquake Engineering & Structural Dynamics 44 (14) (2015) 2377–2399.

[16] R. Sainct, C. Feau, J.-M. Martinez, J. Garnier, Efficient methodology for seismic fragility curves estimation by active learning on support vector machines, Structural Safety 86 (2020) 101972.

[17] C. Mai, M. Spiridonakos, E. Chatzi, B. Sudret, Surrogate modeling for stochastic dynamical systems by combining nonlinear autoregressive with exogenous input models and polynomial chaos expansions, Int. J. Uncertainty Quant. 6 (2016) 313–339.

[18] I. Zentner, Numerical computation of fragility curves for NPP equipment, Nuclear Engineering and Design 240 (6) (2010) 1614–1621. doi:https://doi.org/10.1016/j.nucengdes.2010.02.030.

[19] P. Gehl, J. Douglas, D. M. Seyedi, Influence of the number of dynamic analyses on the accuracy of structural response estimates, Earthquake Spectra 31 (1) (2015) 97–113. doi:https://doi.org/10.1193/102912EQS320M.

[20] K. Trevlopoulos, C. Feau, I. Zentner, Parametric models averaging for optimized non-parametric fragility curve estimation based on intensity measure data clustering, Structural Safety 81 (2019) 101865.

[21] W. Chu, M. Zinkevich, L. Li, A. Thomas, B. Tseng, Unbiased online active learning in data streams, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 195–203.

[22] B. Delyon, F. Portier, Asymptotic optimality of adaptive importance sampling, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, Curran Associates Inc., Red Hook, NY, USA, 2018, p. 3138–3148.

[23] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning: data mining, inference and prediction, Springer, 2009.

[24] C. Gong, W. Zhou, Importance sampling-based system reliability analysis of corroding pipelines considering multiple failure modes, Reliability Engineering & System Safety 169 (2018) 199 – 208.

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

[26] S. Rezaeian, A. Der Kiureghian, Simulation of synthetic ground motions for specified earthquake and site characteristics, Earthquake Engineering & Structural Dynamics 39 (10) (2010) 1155–1180.

[27] C. Mai, K. Konakli, B. Sudret, Seismic fragility curves for structures using non-parametric representations, Frontiers of Structural and Civil Engineering 11 (2) (2017) 169–186.

[28] N. Kwong, A. Chopra, Evaluation of the exact conditional spectrum and generalized conditional intensity measure methods for ground motion selection, Earthquake Engineering & Structural Dynamics 45 (2015) 757–777.

[29] N. Ambraseys, P. Smit, R. Berardi, D. Rinaldis, F. Cotton, C. Berge, Dissemination of european strongmotion data, cD-ROM collection. European Commission, Directorate-General XII, Environmental and Climate Programme, ENV4-CT97-0397, Brussels, Belgium (2000).

[30] I. Zentner, A general framework for the estimation of analytical fragility functions based on multivariate probability distributions, Structural Safety 64 (2017) 54–61. doi:https://doi.org/10.1016/j.strusafe.2016.09.003.
URL http://www.sciencedirect.com/science/article/pii/S016747301630090X

[31] M. Hariri-Ardebili, V. Saouma, Probabilistic seismic demand model and optimal intensity measure for concrete dams, Structural Safety 59 (2016) 67–85. doi:https://doi.org/10.1016/j.strusafe.2015.12.001.
URL http://www.sciencedirect.com/science/article/pii/S0167473015000934

[32] M. Ciano, M. Gioffrè, M. Grigoriu, The role of intensity measures on the accuracy of seismic fragilities, Probabilistic Engineering Mechanics 60 (2020) 103041.

[33] N. Luco, C. A. Cornell, Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions, Earthquake Spectra 23 (2) (2007) 357–392. arXiv:https://doi.org/10.1193/1.2723158, doi:10.1193/1.2723158.
URL https://doi.org/10.1193/1.2723158

[34] J. E. Padgett, B. G. Nielson, R. DesRoches, Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios, Earthquake Engineering & Structural Dynamics 37 (5) (2008) 711–725. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.782, doi:10.1002/eqe.782.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.782

[35] G. E. P. Box, D. R. Cox, An analysis of transformations, J. R. Stat. Soc. B 26 (1964) 211–252.

[36] I. Zentner, M. Gündel, N. Bonfils, Fragility analysis methods: Review of existing approaches and application, Nuclear Engineering and Design 323 (2017) 245 – 258.

[37] D. Lallemant, A. Kiremidjian, H. Burton, Statistical procedures for developing earthquake damage fragility curves, Earthquake Engineering & Structural Dynamics 44 (9) (2015) 1373–1389.

[38] M. M. Zuniga, A. Murangira, T. Perdrizet, Structural reliability assessment through surrogate based importance sampling with dimension reduction, Reliability Engineering & System Safety 207 (2021) 107289.

[39] W. Chu, M. Zinkevich, L. Li, A. Thomas, B. Tseng, Unbiased online active learning in data streams, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD

'11, Association for Computing Machinery, New York, NY, USA, 2011, p. 195–203.

[40] G. Box, G. Tiao, Bayesian Inference in Statistical Analysis, Addison-Wesley, 1973.

[41] M. Keller, A. Popelin, N. Bousquet, E. Remy, Nonparametric estimation of the probability of detection of flaws in an industrial component, from destructive and nondestructive testing data, using approximate bayesian computation, Risk Analysis 35 (2015) 1595–1610.

[42] A. Owen, Y. Zhou, Safe and effective importance sampling, Journal of the American Statistical Association 95 (449) (2000) 135–143.

[43] T. Hesterberg, Weighted average importance sampling and defensive mixture distributions, Technometrics 37 (2) (1995) 185–194.

[44] B. Echard, N. Gayton, M. Lemaire, N. Relun, A combined importance sampling and kriging reliability method for small failure probabilities with time-demanding numerical models, Reliability Engineering & System Safety 111 (2013) 232 – 240.

[45] V. Chabridon, M. Balesdent, J.-M. Bourinet, J. Morio, N. Gayton, Evaluation of failure probability under parameter epistemic uncertainty: application to aerospace system reliability assessment, Aerospace Science and Technology 69 (2017) 526–537.

[46] J. Morio, M. Balesdent, Estimation of rare event probabilities in complex aerospace and other systems: a practical approach, Woodhead Publishing, 2015.

[47] F. Touboul, P. Sollogoub, N. Blay, Seismic behaviour of piping systems with and without defects: experimental and numerical evaluations, Nuclear Engineering and Design 192 (2) (1999) 243 – 260.

[48] T. Charras, J. Kichenind, Développer dans cast3m (2011).
URL http://www-cast3m.cea.fr/

[49] F. Touboul, N. Blay, P. Sollogoub, S. Chapuliot, Enhanced seismic criteria for piping, Nuclear Engineering and Design 236 (1) (2006) 1 – 9.

[50] A. W. v. d. Vaart, Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.

[51] P. Hall, C. Heyde, Z. Birnbaum, E. Lukacs, Martingale Limit Theory and Its Application, Communication and Behavior, Elsevier Science, 2014.