Estimation de courbes de fragilité sismique par planification séquentielle d'expériences.

Clément Gauchy [12] & Cyril Feau [1] & Josselin Garnier [2]

[1] *Universite Paris-Saclay, CEA, Service d'Études Mécaniques et Thermiques, 91191, Gif-sur-Yvette, France, E-mail: clement.gauchy@cea.fr, cyril.feau@cea.fr*
[2] *CMAP, Ecole Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau Cedex, France, E-mail: josselin.garnier@polytechnique.edu*

Résumé. Les études probabilistes de sûreté sismique consistent à évaluer les probabilités de défaillance de structures mécaniques soumises à des excitations sismiques. Ces études nécessitent l'estimation de courbes de fragilité sismique, qui sont la probabilité de défaillance de la structure conditionnellement à une mesure d'intensité du signal sismique. Cependant, leur estimation requiert de nombreuses expériences numériques qui peuvent être très coûteuses en temps de calcul, ce qui rend l'estimation par une méthode Monte Carlo inappropriée. Nous proposons dans ce papier de construire un algorithme de planification séquentielle d'expériences en supposant un *a priori* de processus Gaussien sur la réponse du code de calcul mécanique.

Mots-clés. processus Gaussien, planification séquentielle, courbes de fragilité.

Abstract. Seismic probabilistic risk assessment studies consist in evaluating the probabilities of failure of mechanical structures when submitted to seismic ground motions. These studies are often concentrated on fragility curve estimation. The fragility curve is the probability of failure of the structure conditionally to a seismic intensity measure. However, its estimation requires computer experiments involving huge computation time. Such a computational burden makes crude Monte Carlo methods untractable, fragility curves estimation must then be economical in terms of sample size. We propose an algorithm of sequential planning of experiments by supposing a Gaussian process *prior* on the output of the mechanical computer model.

Keywords. Gaussian process, sequential planning of experiments, fragility curves.

1 Introduction

The estimation of so-called fragility curves is a crucial part of seismic probabilistic risk assessment (SPRA) or probabilistic based earthquake engineering (PBEE). The fragility curve is a way to evaluate structural reliability using a sample of seismic ground motions, which are characterized by their intensity measure (IM) (e.g maximum acceleration of the seimic signal given a time frame). It consists in the conditional probability $\mathbb{P}(Y > C|\text{IM} = \text{a})$ that a specific mechanical demand Y exceeds a threshold of acceptable structural behaviour C for a given seismic intensity of value a. The mechanical demand Y often needs time consuming computation using complex computer codes, making the estimation of fragility curves intractable. This article proposes a Bayesian algorithm of sequential design of experiments, inspired from Bayesian algorithms of the same kind for probability of failure estimation [1] or quantile estimation [2]. The proposed Bayesian algorithm for sequential design of experiments is presented in Section 2. The Section 3 details the practical implementation of the algorithm. Finally, Section 4 presents the performance of the algorithm on a real industrial application concerning a simplified part of a piping system of a French Pressurized Water Reactor.

2 Bayesian decision theory framework

Let $\mathcal{X}$ be a compact set of $\mathbb{R}$ and X an $\mathcal{X}$-valued random variable, in SPRA studies $X = \log \mathrm{IM}$. We define the following nonparametric regression model linking the seismic intensity $x \in \mathcal{X}$ to the log structural mechanical demand $y(x)$:

$$y(x) = g(x) + \varepsilon \ , \tag{1}$$

where ε is a Gaussian homoscedastic noise such that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The usual regression model used in SPRA studies is the log linear model [3], we keep the homoscedasticity assumption but we propose a non-parametric regression function g instead of a linear regression. The problem to be considered here is to estimate the *fragility curve* expressed as

$$\Psi(x; g) = \mathbb{P}_\varepsilon \left(y(X) > C | X = x \right) = \Phi \left(\frac{g(x) - C}{\sigma_\varepsilon} \right) \ , \tag{2}$$

which corresponds to the probability of exceedance of the level C given an input parameter x. The noise variance σ_ε^2 is supposed known while g is unknown. Based on a sequential budget of evaluation points $\mathcal{D}_N = (X_i)_{1 \leq i \leq N}$ with corresponding observed values $(y(X_i))_{1 \leq i \leq N}$, it is possible to provide an estimator $\widehat{\Psi}_N$ of Ψ. After choosing an estimator $\widehat{\Psi}_N$, one has to focus on the good choices of a sequential strategy of evaluation $(X_i)_{1 \leq i \leq N}$. For that matter, we follow the lines of [1] to propose a Bayesian decision-theoretic framework to solve the decision problem of choosing $\mathcal{D}_N$. Given a loss function $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$, we measure the quality of the design of experiments $\mathcal{D}_N$ by the approximation error $e(\mathcal{D}_N, g) = \ell(\Psi, \widehat{\Psi}_N)$. In this paper we consider the Integrated Squared Error (ISE):

$$e(\mathcal{D}_N, g) = \int_{\mathcal{X}} (\Psi(t; g) - \widehat{\Psi}_N(t))^2 \eta(t) dt \ , \tag{3}$$

where η is the pdf of a probability measure with respect to the Lebesgue measure on $\mathcal{X}$. We consider a Bayesian framework to model the uncertainty about the regression function g, we assume that g is a realization of a real-valued Gaussian process G defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P}_0)$. We define the random observation tainted by the Gaussian process uncertainty

$$Y(x) = G(x) + \varepsilon \ . \tag{4}$$

Denote by $\mathcal{F}_n$ the σ-algebra generated by $(X_i, Y(X_i))_{1\leq i\leq n}$ for $1 \leq n \leq N$. Given a Gaussian process G and a set of evaluation points $\mathcal{D}_N$, the optimal estimator $\widehat{\Psi}_n$ that minimizes $\mathbb{E}_0\left[\int_a^b (\Psi(t;G) - \widehat{\Psi}_n(t))^2\eta(t)dt \middle| \mathcal{F}_n\right]$ among the $\mathcal{F}_n$ - measurable estimators is:

$$\widehat{\Psi}_n(t) = \mathbb{E}_0\left[\Psi(t;G)\middle|\mathcal{F}_n\right] = \Phi\left(\frac{\widehat{G}_n(t) - C}{\sigma_n(t)}\right) \ , \tag{5}$$

where $\sigma_n(x)^2 = \widehat{\sigma}_n(x)^2 + \sigma_\varepsilon^2$, given that $(G(t)|\mathcal{F}_n) \sim \mathcal{N}(\widehat{G}_n(t), \widehat{\sigma}_n(t)^2)$ using kriging equations. We choose the optimal strategy $\mathcal{D}_N^* = (X_i^*)_{1\leq i\leq N}$ as the minimizer of the *Bayes risk*:

$$\mathcal{D}_N^* = \underset{\mathcal{D}_N \in \mathcal{X}^N}{\operatorname{argmin}} \, \mathbb{E}_0[e(\mathcal{D}_N, G)] \ , \tag{6}$$

where $\mathbb{E}_0$ is the expectation with respect to $\mathbb{P}_0$. The optimal strategy can be formally obtained. We define $R_N = \mathbb{E}_0\left[\int_a^b (\Psi(t;G) - \widehat{\Psi}_N(t))^2\eta(t)dt \middle| \mathcal{F}_N\right]$ the terminal risk and using backward induction:

$$R_n = \min_{x\in\mathcal{X}} \mathbb{E}_0\left[R_{n+1}|X_{n+1} = x, \mathcal{F}_n\right] \ , \tag{7}$$

$$X_{n+1}^* = \underset{x\in\mathcal{X}}{\operatorname{argmin}} \, R_n \ , \tag{8}$$

for $0 \leq n \leq N-1$. However, the optimal evaluation points for a finite horizon N suffer from the curse of dimensionality. So, in the rest of the paper, we build a sub-optimal set of evaluation points by resolving the optimization problem (8) for $N = 1$. We thus define a $\mathcal{F}_n$-measurable *sampling criterion* $J_n(x)$:

$$J_n(x) = \mathbb{E}_0\left[\int_{\mathcal{X}} (\Psi(t;G) - \widehat{\Psi}_{n+1}(t))^2\eta(t)dt \middle| X_{n+1} = x, \mathcal{F}_n\right] \ , \tag{9}$$

We define the *Stepwise Uncertainty Reduction* (SUR) - based set of evaluation points $\mathcal{D}_N^{\text{SUR}} = (X_i)_{1\leq i\leq N}$ such that for $1 \leq n \leq N-1$:

$$X_{n+1} = \underset{x\in\mathcal{X}}{\operatorname{argmin}} \, J_n(x) \ , \tag{10}$$

3 Practical computation of the sampling criterion

This section is devoted to derive a numerical approximation of the sampling criterion $J_n(x)$ defined in Equation 9. First of all remark that

$$\begin{aligned} J_n(x) &= \int_{\mathcal{X}} \mathbb{E}_0\left[\mathbb{E}_0\left[\left(\Psi(t;G)-\widehat{\Psi}_{n+1}(t)\right)^2 \middle| \mathcal{F}_{n+1}\right] \middle| X_{n+1}=x,\mathcal{F}_n\right]\eta(t)dt \\ &= \int_{\mathcal{X}} \mathbb{E}_0\left[\mathbb{E}_0\left[\Psi(t;G)^2-\widehat{\Psi}_{n+1}(t)^2 \middle| \mathcal{F}_{n+1}\right] \middle| X_{n+1}=x,\mathcal{F}_n\right]\eta(t)dt \\ &= \int_{\mathcal{X}} \mathbb{E}_0\left[\Psi(t;G)^2 \middle| X_{n+1}=x,\mathcal{F}_n\right] - \mathbb{E}_0\left[\widehat{\Psi}_{n+1}(t)^2 \middle| X_{n+1}=x,\mathcal{F}_n\right]\eta(t)dt \end{aligned}$$

We can now precise the two expectations inside the integrands: For the first term, we have

$$\mathbb{E}_0\left[\Psi(t;G)^2 \middle| X_{n+1}=x,\mathcal{F}_n\right] = \mathbb{E}_{Z\sim\mathcal{N}(\widehat{G}_n(t),\widehat{\sigma}_n(t)^2)}\left[\Phi\left(\frac{Z-C}{\sigma_\varepsilon}\right)^2\right] \quad (11)$$

For the second term, we have

$$\mathbb{E}_0\left[\widehat{\Psi}_{n+1}(t)^2 \middle| X_{n+1}=x,\mathcal{F}_n\right] = \mathbb{E}_{Z\sim\mathcal{N}(\widehat{G}_n(x),\widehat{\sigma}_n(x)^2)}\left[\Phi\left(\frac{\widehat{G}_{n+1}(t;Z)-C}{\sigma_{n+1}(t)}\right)^2\right] \quad (12)$$

where $\widehat{G}_{n+1}(t;Z)$ and $\widehat{\sigma}_{n+1}(t)^2$ are respectively the conditional mean and variance of the GP G knowing the observations $(X_i, Y(X_i))_{1\le i\le n}$ and the virtual output Z at point $X_{n+1}=x$. Remark that $\widehat{G}_{n+1}$ is an affine function in Z and $\widehat{\sigma}_{n+1}$ depends only on $(X_i)_{1\le i\le n+1}$. The two expectations can be approximated using Gauss-Hermite quadrature (for the heat kernel e^{-x^2}) with Q points with quadrature points $(u_q)_{1\le q\le Q}$ and quadrature weights $(\omega_q)_{1\le i\le Q}$. The integral with respect to η can be approximated by a Riemann sum with regular grid $(t_i)_{1\le i\le T}$ of interval length ΔT (Gaussian quadrature could be used as well):

$$\left\{\begin{array}{ll} J_n(x) & \approx \frac{1}{\sqrt{\pi}}\sum_{i=1}^{T}\sum_{q=1}^{Q}\Delta T\omega_q\left(\Phi\left(\frac{z_{n,q}(t_i)-C}{\sigma_\varepsilon}\right)^2-\Phi\left(\frac{\widehat{G}_{n+1}(t_i;z_{n,q}(x))-C}{\sigma_{n+1}(t_i)}\right)^2\right)\eta(t_i) \\ z_{n,q}(x) & = \widehat{G}_n(x)+\widehat{\sigma}_n(x)u_q\sqrt{2} \end{array}\right. \quad (13)$$

4 Industrial application: safety water pipe of a French Pressurized Reactor

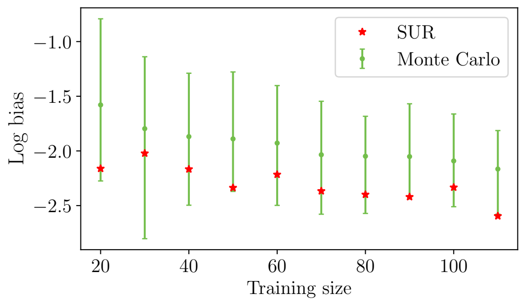
The following test case corresponds to a piping system which is a simplified part of a secondary line of a French Pressurized Water Reactor [4]. The structural mechanical demand in this case is the out-of-plane rotation R of a specific elbow. A single degree of freedom nonlinear oscillator is calibrated on the Finite Element Model of the pipe implemented in CAST3M [5] and is used to validate the proposed sequential planning of experiments. The synthetic seismic signals dataset – developed and used in [6] – is filtered by a fictitious linear single-mode building at 5 Hz and damped at 2 %. 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 (first eigenfrequency of the pipe) and 1 % damping ratio. The rotation threshold considered here is $C = 1.8°$. This is the 90 %-level quantile of the full dataset of simulations based on 10^5 synthetic seismic ground motions. The Gaussian process used has a zero mean function and stationary Matèrn 5/2 covariance function. The performance of the proposed sequential planning strategy is evaluated by comparing the estimator $\widehat{\Psi}_n$ with a non-parametric estimation of the fragility curve Ψ_{ref} with the full dataset of 10^5 synthetic signals (see e.g. [6]). We compute the *bias* b_n and the *posterior variance* v_n of a design $\mathcal{D}_n$ by:

$$b_n = \int (\widehat{\Psi}_n(t) - \Psi_{\text{ref}}(t))^2 \eta(t) dt \ , \quad v_n = \mathbb{E}_0 \left[\int (\Psi(t; G) - \widehat{\Psi}_n(t))^2 \eta(t) dt \Big| \mathcal{F}_n \right] \ . \qquad (14)$$

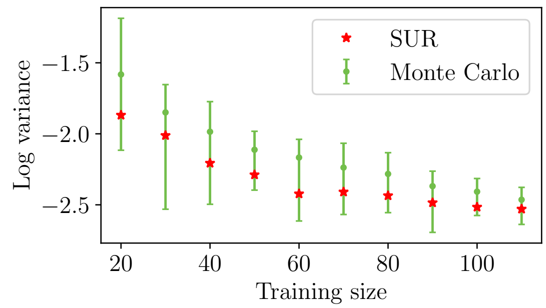
For this case we choose $\eta(t) = \mathbb{1}_{[a,b]}(t)$ ($a = 3,1\ m/s^2,\ b = 31,2\ m/s^2$). The SUR procedure starts by choosing 10 synthetic seismic ground motions in a stratified manner, by dividing the full dataset into 10 partitions of same size using the $(j/10)_{1 \le j \le 9}$-level quantiles of the seismic intensity measure. They are the first points of the design set and are used to optimize the Gaussian process hyperparameters (including σ_ε). After that, we propose a brute force algorithm to solve the optimization problem (10). A set of candidate seismic signals with log intensity measure $(U_i)_{1 \le i \le\ m}$ ($m = 200$) are chosen in the dataset of 10^5 synthetic seismic ground motions. The candidate set is built by using the same partition into 10 subdomains as for initialization. 20 seismic signals are chosen at random for each subdomain. We then evaluate the sampling criterion using Eq. (13) for each seismic signal in the set of candidates. We then take the seismic signal that minimizes the criterion $i^* = \text{argmin}_{1 \le i \le\ m} J_{n-1}(U_i), \quad X_n = U_{i^*}$ and then we add X_n to our design $\mathcal{D}_n$ of size n and compute the log mechanical demand $y(X_n)$ using our computer code. The Gaussian process hyperparameters are optimized every 10 iterations.

Figure 1 compares the biases b_n and variances v_n as a function of the training size n for the SUR strategy and for 100 replications of Monte-Carlo sampling in the 10^5-sized dataset of seismic signals. The SUR strategy offers good performance in terms of bias between the estimated fragility curve $\widehat{\Psi}_n$ and the reference fragility curve Ψ_{ref} and is equivalent to Monte-Carlo sampling in terms of variance. These results motivate further research in this topic.

Acknowledgments: Clément Gauchy would like to thanks Vincent Chabridon and Elias Fekhari for fruitful discussions about Bayesian decision theory.



(a) Log bias



(b) Log variance

Figure 1: Comparison of the bias b_n and variance v_n (in log) between Monte-Carlo sampling and our SUR strategy. The green error bars correspond to the interval between the 10% and 90% level quantile of 100 replications of Monte Carlo sampling.

References

[1] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, April 2011. doi: 10.1007/s11222-011-9241-4.

[2] Aurélie Arnaud, Julien Bect, Mathieu Couplet, Alberto Pasanisi, and Emmanuel Vazquez. Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences. *42èmes Journées de Statistique*, 2010.

[3] Bruce R. Ellingwood and Kursat Kinali. Quantifying and communicating uncertainty in seismic risk assessment. *Structural Safety*, 31(2):179–187, 2009. doi: https://doi.org/10.1016/j.strusafe.2008.06.001. Risk Acceptance and Risk Communication.

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

[5] URL `http://www-cast3m.cea.fr/`.

[6] Rémi Sainct, Cyril Feau, Jean-Marc Martinez, and Josselin Garnier. Efficient methodology for seismic fragility curves estimation by active learning on support vector machines. *Structural Safety*, 86:101972, 2020.