ETICS 2022

Estimation of seismic fragility curves by sequential design of experiments

Clément GAUCHY (CEA, CMAP), Cyril FEAU (CEA), Josselin GARNIER (CMAP)

Seismic safety studies

Probabilistic safety studies in earthquake engineering

Probabilistic safety studies aim to evaluate the reliability of a mechanical structure subjected to seismic hazard, they are broken down into three steps:

- **1)** The estimation of an annual occurrence probability of a seismic excitation of specific intensity

Probabilistic safety studies in earthquake engineering

Probabilistic safety studies aim to evaluate the reliability of a mechanical structure subjected to seismic hazard, they are broken down into three steps:

- **1)** The estimation of an annual occurrence probability of a seismic excitation of specific intensity
- **2)** The estimation of the probability of failure of a structure conditional to the seismic intensity (seismic fragility curve)

Probabilistic safety studies in earthquake engineering

Probabilistic safety studies aim to evaluate the reliability of a mechanical structure subjected to seismic hazard, they are broken down into three steps:

- **1)** The estimation of an annual occurrence probability of a seismic excitation of specific intensity
- **2)** The estimation of the probability of failure of a structure conditional to the seismic intensity (seismic fragility curve)
- **3)** The evaluation of the annual failure probability of the structure, evaluated thanks to the 2 steps above

↪ This presentation will focus on the step **2)**.

Schematic representation

Mechanical parameters

$\mathbf{x} \in \mathcal{D}_{\mathbf{x}} \subset \mathbb{R}^d$

$\hookrightarrow$ Young modulus, boundary conditions

Computer code of the structure

$\hookrightarrow$ Piping system of a nuclear power plant.

Numerical simulation output

$z(a, \mathbf{x}) \in \mathbb{R}$

$\hookrightarrow$ Out of plane rotation of a pipe elbow

Seismic ground motion

$a \in \mathcal{D}_a \subset \mathbb{R}$

$\hookrightarrow$ Spectral acceleration.

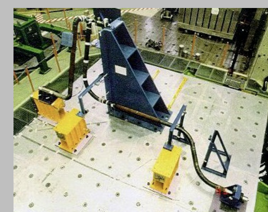
Schematic representation

Mechanical parameters

$\mathbf{x} \in \mathcal{D}_{\mathbf{x}} \subset \mathbb{R}^d$

$\hookrightarrow$ Young modulus, boundary conditions

Computer code of the structure



$\hookrightarrow$ Piping system of a nuclear power plant.

Numerical simulation output

$z(a, \mathbf{x}) \in \mathbb{R}$

$\hookrightarrow$ Out of plane rotation of a pipe elbow

Seismic ground motion

$a \in \mathcal{D}_a \subset \mathbb{R}$

$\hookrightarrow$ Spectral acceleration.

The output of the simulation $z(a, \mathrm{x})$ is stochastic (i.e. for a same value of (a, x) the value of the output can change)

Seismic intensity measurement

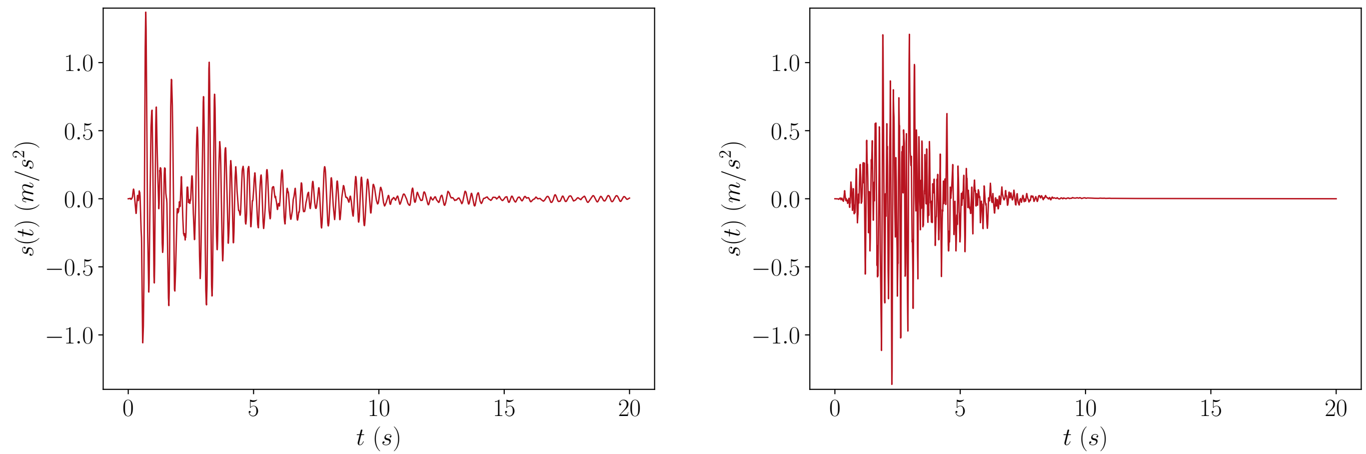


Figure: Two earthquake accelerograms with the same peak ground acceleration

Generally a measure of seismic intensity is a scalar quantity derived from the time signal. Example: the peak ground acceleration of a seismic ground motion with accelerogram $t \to s(t)$.

$$a = \max_{t \in [0,T]} |s(t)| .$$

$\hookrightarrow$ There is no uniqueness between a seismic measurement intensity value and a seismic signal.

Seismic fragility curve

The fragility curve is the conditional probability of failure of the structure:

$$\Psi(a, \mathbf{x}) = \mathbb{P}\left(z(A, \mathbf{X}) > C | A = a, \mathbf{X} = \mathbf{x}\right)$$

Seismic fragility curve

The fragility curve is the conditional probability of failure of the structure:

$$\boldsymbol{\Psi}(\boldsymbol{a}, \mathbf{x}) = \mathbb{P}\left(\boldsymbol{z}(\boldsymbol{A}, \mathbf{X}) > \boldsymbol{C} | \boldsymbol{A} = \boldsymbol{a}, \mathbf{X} = \mathbf{x}\right)$$

$\boldsymbol{A}$: Scalar variable representing the intensity of the seismic solicitation.

Seismic fragility curve

The fragility curve is the conditional probability of failure of the structure:

$$\boldsymbol{\Psi}(\boldsymbol{a}, \mathrm{x}) = \mathbb{P}\left(\boldsymbol{z}(\boldsymbol{A}, \mathrm{X}) > \boldsymbol{C} | \boldsymbol{A} = \boldsymbol{a}, \mathrm{X} = \mathrm{x}\right)$$

$\boldsymbol{A}$: Scalar variable representing the intensity of the seismic solicitation

X: Mechanical parameters of the structure.

Seismic fragility curve

The fragility curve is the conditional probability of failure of the structure:

$$\Psi(a, \mathrm{x}) = \mathbb{P}\left(z(A, \mathrm{X}) > C | A = a, \mathrm{X} = \mathrm{x}\right)$$

A: Scalar variable representing the intensity of the seismic solicitation

X: Mechanical parameters of the structure.

C: Critical level above which the structure is considered in failure state.

Seismic fragility curve

The fragility curve is the conditional probability of failure of the structure:

$$\Psi(\boldsymbol{a}, \mathbf{x}) = \mathbb{P}\left(\boldsymbol{z}(\boldsymbol{A}, \mathrm{X}) > \boldsymbol{C} | \boldsymbol{A} = \boldsymbol{a}, \mathrm{X} = \mathrm{x}\right)$$

$\boldsymbol{A}$: Scalar variable representing the intensity of the seismic solicitation

$\mathbf{X}$: Mechanical parameters of the structure.

$\boldsymbol{C}$: Critical level above which the structure is considered in failure state.

$\boldsymbol{z}$: Computer model output

Scientific goals

We would like to study the impact of the uncertainty on the vector of mechanical parameters x of the structure on the fragility curves $\boldsymbol{a} \to \boldsymbol{\Psi}(\boldsymbol{a}, \mathrm{x})$.

Scientific goals

We would like to study the impact of the uncertainty on the vector of mechanical parameters x of the structure on the fragility curves $\boldsymbol{a} \rightarrow \boldsymbol{\Psi}(\boldsymbol{a}, \mathrm{x})$.

We model this vector by a random variable $\mathbf{X}$ following a probability distribution:

$$\mathbf{X} \sim \mathbb{P}_{\mathbf{X}}$$

Scientific goals

We would like to study the impact of the uncertainty on the vector of mechanical parameters x of the structure on the fragility curves $\boldsymbol{a} \to \boldsymbol{\Psi}(\boldsymbol{a}, \mathrm{x})$.

We model this vector by a random variable $\mathbf{X}$ following a probability distribution:

$$\mathbf{X} \sim \mathbb{P}_{\mathrm{X}}$$

With a Monte Carlo sample $(\mathbf{X}_i)_{1\leq i\leq M}$ such that $\mathbf{X}_i \sim \mathbb{P}_{\mathrm{X}}$ we can propagate the uncertainty on the fragility curve by studying the distribution of $(\boldsymbol{a} \to \boldsymbol{\Psi}(\boldsymbol{a}, \mathbf{X}_i))_{1\leq i\leq M}$.

Scientific goals

We would like to study the impact of the uncertainty on the vector of mechanical parameters x of the structure on the fragility curves $\boldsymbol{a} \to \boldsymbol{\Psi}(\boldsymbol{a}, \mathrm{x})$.

We model this vector by a random variable $\mathbf{X}$ following a probability distribution:

$$\mathbf{X} \sim \mathbb{P}_{\mathrm{X}}$$

With a Monte Carlo sample $(\mathbf{X}_i)_{1 \leq i \leq M}$ such that $\mathbf{X}_i \sim \mathbb{P}_{\mathrm{X}}$ we can propagate the uncertainty on the fragility curve by studying the distribution of $(\boldsymbol{a} \to \boldsymbol{\Psi}(\boldsymbol{a}, \mathbf{X}_i))_{1 \leq i \leq M}$.

Problem: About $\mathbf{10^4}$ simulations to estimate a curve $\boldsymbol{a} \to \boldsymbol{\Psi}(\boldsymbol{a}, \mathbf{x})$ for a $\mathbf{x}$ fixed in the classical way. For $\boldsymbol{M} = \mathbf{1000}$ it would be necessary to do 10^7 simulations.

10^7 CAST3M simulations $\approx$ 100 days of computation time.

Goal: Provide a best estimate of $\Psi(a, \mathrm{x})$ with a fixed budget of CAST3M simulations.

Sequential design of experiments

Statistical model

We must therefore propose a statistical model on the output of the CAST3M simulation $\boldsymbol{z}(\boldsymbol{a}, \mathrm{x})$:

Statistical model

We must therefore propose a statistical model on the output of the CAST3M simulation $z(a, \mathrm{x})$:

$$y(a, \mathrm{x}) = g(a, \mathrm{x}) + \varepsilon ,$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $y(a, \mathrm{x}) = \log(z(a, \mathrm{x}))$.

The fragility curve with this model writes:

$$\Psi(a, x; g) = \Phi\left(\frac{g(a, x) - \log(C)}{\sigma_\varepsilon}\right) ,$$

where Φ is the cdf of the standard Gaussian distribution

Gaussian process

We define a Gaussian process $\boldsymbol{G} : \mathrm{x} \to \boldsymbol{G}(\mathrm{x})$ thanks to the Gaussian vectors. Let $(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(p)})$, the process $\boldsymbol{G}$ is said to be Gaussian if the vector $(\boldsymbol{G}(\mathrm{x}^{(1)}), \ldots, \boldsymbol{G}(\mathrm{x}^{(p)}))$ is Gaussian.

Gaussian process

We define a Gaussian process $\boldsymbol{G} : \mathrm{x} \to \boldsymbol{G}(\mathrm{x})$ thanks to the Gaussian vectors. Let $(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(p)})$, the process $\boldsymbol{G}$ is said to be Gaussian if the vector $(\boldsymbol{G}(\mathrm{x}^{(1)}), \ldots, \boldsymbol{G}(\mathrm{x}^{(p)}))$ is Gaussian.

A Gaussian process is entirely defined by its mean function:

$$\boldsymbol{m}(\mathrm{x}) = \mathbb{E}[\boldsymbol{G}(\mathrm{x})] \ ,$$

and its covariance function:

$$\boldsymbol{\Sigma}(\mathrm{x}, \tilde{\mathrm{x}}) = \mathbb{E}[(\boldsymbol{G}(\mathrm{x}) - \boldsymbol{\mu}(\mathrm{x}))(\boldsymbol{G}(\tilde{\mathrm{x}}) - \boldsymbol{\mu}(\tilde{\mathrm{x}}))].$$

Gaussian process

We define a Gaussian process $\boldsymbol{G} : \mathrm{x} \to \boldsymbol{G}(\mathrm{x})$ thanks to the Gaussian vectors. Let $(\mathrm{x}^{(1)}, \ldots, \mathrm{x}^{(p)})$, the process $\boldsymbol{G}$ is said to be Gaussian if the vector $(\boldsymbol{G}(\mathrm{x}^{(1)}), \ldots, \boldsymbol{G}(\mathrm{x}^{(p)}))$ is Gaussian.

A Gaussian process is entirely defined by its mean function:

$$\boldsymbol{m}(\mathrm{x}) = \mathbb{E}[\boldsymbol{G}(\mathrm{x})] \, ,$$

and its covariance function:

$$\boldsymbol{\Sigma}(\mathrm{x}, \tilde{\mathrm{x}}) = \mathbb{E}[(\boldsymbol{G}(\mathrm{x}) - \boldsymbol{\mu}(\mathrm{x}))(\boldsymbol{G}(\tilde{\mathrm{x}}) - \boldsymbol{\mu}(\tilde{\mathrm{x}}))].$$

$$\boldsymbol{Y} \sim \mathcal{GP}(\boldsymbol{m}, \Sigma) \, .$$

The Gaussian vector $(\boldsymbol{G}(\mathrm{x}^{(1)}), \ldots, \boldsymbol{G}(\mathrm{x}^{(p)}))$ has mean vector $\boldsymbol{\mu} = (\boldsymbol{m}(\mathrm{x}^{(i)}))_{1 \leq i \leq p}$ and covariance matrix $\boldsymbol{K} = (\Sigma(\mathrm{x}^{(i)}, \mathrm{x}^{(j)}))_{1 \leq i,j \leq p}$.

A Bayesian model

We will model an uncertainty on $\boldsymbol{g}$ by a Gaussian process prior $\boldsymbol{G}$ defined on some probabilistic space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{B}}, \mathbb{P}_0)$. The statistical model then becomes:

$$\boldsymbol{Y}(\boldsymbol{a}, \mathrm{x}) = \boldsymbol{G}(\boldsymbol{a}, \mathrm{x}) + \boldsymbol{\varepsilon} \ ,$$

A Bayesian model

We will model an uncertainty on $\boldsymbol{g}$ by a Gaussian process prior G defined on some probabilistic space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{B}}, \mathbb{P}_0)$. The statistical model then becomes:

$$\boldsymbol{Y}(\boldsymbol{a}, \mathrm{x}) = G(\boldsymbol{a}, \mathrm{x}) + \varepsilon \,,$$

We can propose a Bayesian estimator of the fragility curve. Define $\boldsymbol{\mathcal{F}}_n$ the $\boldsymbol{\sigma}$-algebra defined by $(\boldsymbol{A}_i, \mathrm{X}_i, \boldsymbol{y}(\boldsymbol{A}_i, \mathrm{X}_i))_{1\leq i\leq n}$. The *posterior mean* writes:

$$\widehat{\boldsymbol{\Psi}}_n(\boldsymbol{a}, \mathrm{x}) = \mathbb{E}_{\mathbb{P}_0}[\boldsymbol{\Psi}(\boldsymbol{a}, \boldsymbol{x}; G)|\boldsymbol{\mathcal{F}}_n]$$

A Bayesian model

We will model an uncertainty on $\boldsymbol{g}$ by a Gaussian process prior G defined on some probabilistic space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{B}}, \mathbb{P}_0)$. The statistical model then becomes:

$$\boldsymbol{Y}(\boldsymbol{a}, \mathrm{x}) = \boldsymbol{G}(\boldsymbol{a}, \mathrm{x}) + \boldsymbol{\varepsilon} \ ,$$

We can propose a Bayesian estimator of the fragility curve. Define $\boldsymbol{\mathcal{F}}_n$ the $\boldsymbol{\sigma}$-algebra defined by $(\boldsymbol{A}_i, \mathrm{X}_i, \boldsymbol{y}(\boldsymbol{A}_i, \mathrm{X}_i))_{1 \leq i \leq n}$. The *posterior mean* writes:

$$\widehat{\boldsymbol{\Psi}}_n(\boldsymbol{a}, \mathrm{x}) = \mathbb{E}_{\mathbb{P}_0}[\boldsymbol{\Psi}(\boldsymbol{a}, \boldsymbol{x}; \boldsymbol{G}) | \boldsymbol{\mathcal{F}}_n]$$

We have $(\boldsymbol{G}(\boldsymbol{a}, \mathrm{x}) | \boldsymbol{\mathcal{F}}_n) \sim \boldsymbol{\mathcal{N}}(\widehat{\boldsymbol{G}}_n(\boldsymbol{a}, \mathrm{x}), \widehat{\boldsymbol{\sigma}}_n(\boldsymbol{a}, \mathrm{x})^2)$. Thus:

$$\widehat{\boldsymbol{\Psi}}_n(\boldsymbol{a}, \mathrm{x}) = \boldsymbol{\Phi}\left(\frac{\widehat{\boldsymbol{G}}_n(\boldsymbol{a}, \mathrm{x}) - \log(\boldsymbol{C})}{\boldsymbol{\sigma}_n(\boldsymbol{a}, \mathrm{x})}\right) \ ,$$

where $\boldsymbol{\sigma}_n(\boldsymbol{a}, \mathrm{x})^2 = \widehat{\boldsymbol{\sigma}}_n(\boldsymbol{a}, \mathrm{x})^2 + \boldsymbol{\sigma}_{\varepsilon}^2$

A decision problem

For a given $\boldsymbol{n} \geq \mathbf{1}$, we have to choose a design $\boldsymbol{\mathcal{D}}_n = (\boldsymbol{A}_i, \mathrm{X}_i)_{1 \leq i \leq n}$ where we compute the mechanical response $\boldsymbol{y}(\boldsymbol{A}_i, \mathrm{X}_i)$. **How to measure the quality of the design $\mathcal{D}_n$?**

A decision problem

For a given $n \geq 1$, we have to choose a design $\mathcal{D}_n = (A_i, \mathrm{X}_i)_{1\leq i\leq n}$ where we compute the mechanical response $y(A_i, \mathrm{X}_i)$. **How to measure the quality of the design $\mathcal{D}_n$?**

The *Bayes risk* r_B of a Bayesian estimator of the fragility curve $\widehat{\Psi}_n$ based on a design $\mathcal{D}_n = (A_i, \mathrm{X}_i)_{1\leq i\leq n}$ and computer experiments $y(A_i, \mathrm{X}_i)$ is defined by:

$$r_B(\mathcal{D}_n, G) = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\Psi(\alpha, u; G) - \widehat{\Psi}_n(\alpha, u))^2 dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u)\right] .$$

A decision problem

For a given $n \geq 1$, we have to choose a design $\mathcal{D}_n = (A_i, \mathrm{X}_i)_{1 \leq i \leq n}$ where we compute the mechanical response $y(A_i, \mathrm{X}_i)$. **How to measure the quality of the design $\mathcal{D}_n$?**

The *Bayes risk* r_B of a Bayesian estimator of the fragility curve $\widehat{\Psi}_n$ based on a design $\mathcal{D}_n = (A_i, \mathrm{X}_i)_{1 \leq i \leq n}$ and computer experiments $y(A_i, \mathrm{X}_i)$ is defined by:

$$r_B(\mathcal{D}_n, G) = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\Psi(\alpha, u; G) - \widehat{\Psi}_n(\alpha, u))^2 dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u)\right] .$$

The optimal design $\mathcal{D}_n^* = (A_i^*, \mathrm{X}_i^*)_{1 \leq i \leq n}$ is obtained by minimizing the Bayes risk:

$$\mathcal{D}_n^* = \underset{\mathcal{D}_n \in \mathcal{S}_n}{\operatorname{argmin}}\, r_B(\mathcal{D}_n, G) ,$$

where $\mathcal{S}_n$ is the set of all admissible sequential strategies of size n.

A decision problem

What is necessary for a Bayesian decision problem ?

A decision problem

What is necessary for a Bayesian decision problem ?

- A statistical model and a prior $\mathbb{P}_0$ (here a Gaussian process $\boldsymbol{G}$)

A decision problem

What is necessary for a Bayesian decision problem ?

- A statistical model and a prior $\mathbb{P}_0$ (here a Gaussian process $\boldsymbol{G}$)
- A Bayesian estimator of a quantity of interest (here $\widehat{\boldsymbol{\Psi}}_n$) which depends on a design $\boldsymbol{\mathcal{D}}_n$

A decision problem

What is necessary for a Bayesian decision problem ?

- A statistical model and a prior $\mathbb{P}_0$ (here a Gaussian process $\boldsymbol{G}$)
- A Bayesian estimator of a quantity of interest (here $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{n}}$) which depends on a design $\mathcal{D}_n$
- A loss function (here $\int_{\mathcal{A}\times\mathcal{X}}(\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{u}; \boldsymbol{G}) - \widehat{\boldsymbol{\Psi}}_n(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha})\boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u})$)

A decision problem

What is necessary for a Bayesian decision problem ?

- A statistical model and a prior $\mathbb{P}_0$ (here a Gaussian process $\boldsymbol{G}$)
- A Bayesian estimator of a quantity of interest (here $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{n}}$) which depends on a design $\boldsymbol{\mathcal{D}}_{\boldsymbol{n}}$
- A loss function (here $\int_{\mathcal{A}\times\mathcal{X}}(\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{u}; \boldsymbol{G}) - \widehat{\boldsymbol{\Psi}}_{\boldsymbol{n}}(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha}) \boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u})$)

These **3** parts are necessary to define the Bayes risk and hence the optimal design $\boldsymbol{\mathcal{D}}_{\boldsymbol{n}}^*$

Optimal design of experiments

The optimal design $\mathcal{D}_n^*$ can be obtained formally using dynamic programming. Define

$$R_n = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\Psi(\alpha, u; G) - \widehat{\Psi}_n(\alpha, u))^2 dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u) \middle| \mathcal{F}_n\right] ,$$

and by reverse induction

$$R_k = \min_{a,\mathrm{x}\in\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[R_{k+1} \middle| A_{k+1} = a, \mathrm{X}_{k+1} = \mathrm{x}, \mathcal{F}_k\right] ,$$

$$A_{k+1}^*, \mathrm{X}_{k+1}^* = \operatorname*{argmin}_{a,\mathrm{x}\in\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[R_{k+1} \middle| A_{k+1} = a, \mathrm{X}_{k+1} = \mathrm{x}, \mathcal{F}_k\right] .$$

for $0 \leq k \leq n-1$.

Optimal design of experiments

The optimal design $\mathcal{D}_n^*$ can be obtained formally using dynamic programming. Define

$$R_n = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\Psi(\alpha, u; G) - \widehat{\Psi}_n(\alpha, u))^2 dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u) \middle| \mathcal{F}_n\right] ,$$

and by reverse induction

$$R_k = \min_{a,\mathrm{x}\in\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[R_{k+1} \middle| A_{k+1} = a, \mathrm{X}_{k+1} = \mathrm{x}, \mathcal{F}_k\right] ,$$

$$A_{k+1}^*, \mathrm{X}_{k+1}^* = \operatorname*{argmin}_{a,\mathrm{x}\in\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[R_{k+1} \middle| A_{k+1} = a, \mathrm{X}_{k+1} = \mathrm{x}, \mathcal{F}_k\right] .$$

for $0 \leq k \leq n-1$.

The exact optimal design is intractable to obtain in practice.

Stepwise uncertainty reduction (SUR)

The optimal design is intractable to obtain due to nested nonconvex optimization problem.

[1] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. 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

Stepwise uncertainty reduction (SUR)

The optimal design is intractable to obtain due to nested nonconvex optimization problem.

→ Relax the optimization problem and propose a greedy approach (SUR strategy)[1]

[1] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. 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

Stepwise uncertainty reduction (SUR)

The optimal design is intractable to obtain due to nested nonconvex optimization problem.

→ Relax the optimization problem and propose a greedy approach (SUR strategy)[1]

The idea is to stop the previous induction at $n = 1$:

$$\boldsymbol{A}_{n+1}^{\mathrm{SUR}}, \mathrm{X}_{n+1}^{\mathrm{SUR}} = \underset{a,\mathrm{x} \in \mathcal{A} \times \mathcal{X}}{\operatorname{argmin}} J_n(a, \mathrm{x})$$

$$J_n(a, \mathrm{x}) = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\Psi(\alpha, u; G) - \widehat{\Psi}_{n+1}(\alpha, u))^2 dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u) \middle| A_{n+1} = a, \mathrm{X}_{n+1} = \mathrm{x}, \mathcal{F}_n\right]$$

[1] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. 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

The optimal design is intractable to obtain due to nested nonconvex optimization problem.

→ Relax the optimization problem and propose a greedy approach (SUR strategy)[1]

The idea is to stop the previous induction at $\boldsymbol{n = 1}$:

$$\boldsymbol{A}_{n+1}^{\mathrm{SUR}}, \mathbf{X}_{n+1}^{\mathrm{SUR}} = \underset{\boldsymbol{a}, \mathbf{x} \in \mathcal{A} \times \mathcal{X}}{\operatorname{argmin}} \, \boldsymbol{J}_n(\boldsymbol{a}, \mathbf{x})$$

$$\boldsymbol{J}_n(\boldsymbol{a}, \mathbf{x}) = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A} \times \mathcal{X}} (\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{u}; \boldsymbol{G}) - \widehat{\boldsymbol{\Psi}}_{n+1}(\boldsymbol{\alpha}, \boldsymbol{u}))^2 d\boldsymbol{h}(\boldsymbol{\alpha}) d\mathbb{P}_{\mathbf{X}}(\boldsymbol{u}) \middle| \boldsymbol{A}_{n+1} = \boldsymbol{a}, \mathbf{X}_{n+1} = \mathbf{x}, \mathcal{F}_n\right]$$

Remark that $\boldsymbol{J}_n(\boldsymbol{a}, \boldsymbol{x})$ is an expectation w.r.t. $(\boldsymbol{Y}(\boldsymbol{A}_{n+1}, \mathbf{X}_{n+1}) | \mathcal{F}_n)$.

[1] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. 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

Practical computation of J_n

It is possible to rewrite $\boldsymbol{J_n(a,x)}$ to perform a simple Monte-Carlo loop[2]:

$$\boldsymbol{J_n(a,x) = \int_{\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[\Psi(\alpha,u;G)^2\middle|\mathcal{F}_n\right] - \mathbb{E}_{\mathbb{P}_0}\left[\widehat{\Psi}_{n+1}(\alpha,u)^2\middle|A_{n+1}=a, \mathrm{X}_{n+1}=\mathrm{x}, \mathcal{F}_n\right] dh(\alpha)d\mathbb{P}_{\mathrm{X}}(u)}$$

[2] Clement Gauchy, Cyril Feau, and Josselin Garnier. Estimation of seismic fragility curves by sequential design of experiments.
February 2022.
hal-03588974

Practical computation of J_n

First, remark that $\boldsymbol{J_n(a,x)}$ is an expectation w.r.t. $(\boldsymbol{G}(\boldsymbol{A}_{n+1}, \mathbf{X}_{n+1})|\boldsymbol{\mathcal{F}_n})$.

It is possible to rewrite $\boldsymbol{J_n(a,x)}$ to perform a simple Monte-Carlo loop:

$$\boldsymbol{J_n(a,x)} = \int_{\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[\Psi(\alpha, u; G)^2 \middle| \mathcal{F}_n\right] - \mathbb{E}_{\mathbb{P}_0}\left[\widehat{\Psi}_{n+1}(\alpha, u)^2 \middle| \boldsymbol{A}_{n+1} = \boldsymbol{a}, \mathbf{X}_{n+1} = \mathrm{x}, \mathcal{F}_n\right] dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u)$$

$$\mathbb{E}_{\mathbb{P}_0}\left[\Psi(\alpha, u; G)^2 \middle| \mathcal{F}_n\right] = \mathbb{E}_{Z\sim\mathcal{N}(\widehat{G}_n(\alpha,u), \widehat{\sigma}_n(\alpha,u)^2)}\left[\Phi\left(\frac{Z - \log(C)}{\sigma_\varepsilon}\right)^2\right]$$

Practical computation of J_n

First, remark that $J_n(a,x)$ is an expectation w.r.t. $(G(A_{n+1}, \mathrm{X}_{n+1})|\mathcal{F}_n)$.

It is possible to rewrite $J_n(a,x)$ to perform a simple Monte-Carlo loop:

$$J_n(a,x) = \int_{\mathcal{A}\times\mathcal{X}} \mathbb{E}_{\mathbb{P}_0}\left[\Psi(\alpha,u;G)^2 \middle| \mathcal{F}_n\right] - \mathbb{E}_{\mathbb{P}_0}\left[\widehat{\Psi}_{n+1}(\alpha,u)^2 \middle| A_{n+1} = a, \mathrm{X}_{n+1} = \mathrm{x}, \mathcal{F}_n\right] dh(\alpha) d\mathbb{P}_{\mathrm{X}}(u)$$

$$\mathbb{E}_{\mathbb{P}_0}\left[\Psi(\alpha,u;G)^2 \middle| \mathcal{F}_n\right] = \mathbb{E}_{Z\sim\mathcal{N}(\widehat{G}_n(\alpha,u),\widehat{\sigma}_n(\alpha,u)^2)}\left[\Phi\left(\frac{Z-\log(C)}{\sigma_\varepsilon}\right)^2\right]$$

$$\mathbb{E}_{\mathbb{P}_0}\left[\widehat{\Psi}_{n+1}(\alpha,u)^2 \middle| A_{n+1} = a, \mathrm{X}_{n+1} = \mathrm{x}, \mathcal{F}_n\right] = \mathbb{E}_{Z\sim\mathcal{N}(\widehat{G}_n(a,\mathrm{x}),\widehat{\sigma}_n(a,\mathrm{x})^2)}\left[\Phi\left(\frac{\widehat{G}_{n+1}(\alpha,u;Z)-\log(C)}{\sigma_{n+1}(\alpha,u)}\right)^2\right]$$

where $\widehat{G}_{n+1}(\alpha,u;Z)$ is the conditional mean of the GP with "virtual" output Z at design point $(A_{n+1}, \mathrm{X}_{n+1})$.

Practical computation of J_n

The first integral is approximated using Monte-Carlo simulation with a sample $(\boldsymbol{\alpha}_i, \boldsymbol{U}_i)_{1 \leq i \leq N}$ drawn from the product measure $\boldsymbol{h} \otimes \mathbb{P}_{\mathbf{X}}$.

Practical computation of J_n

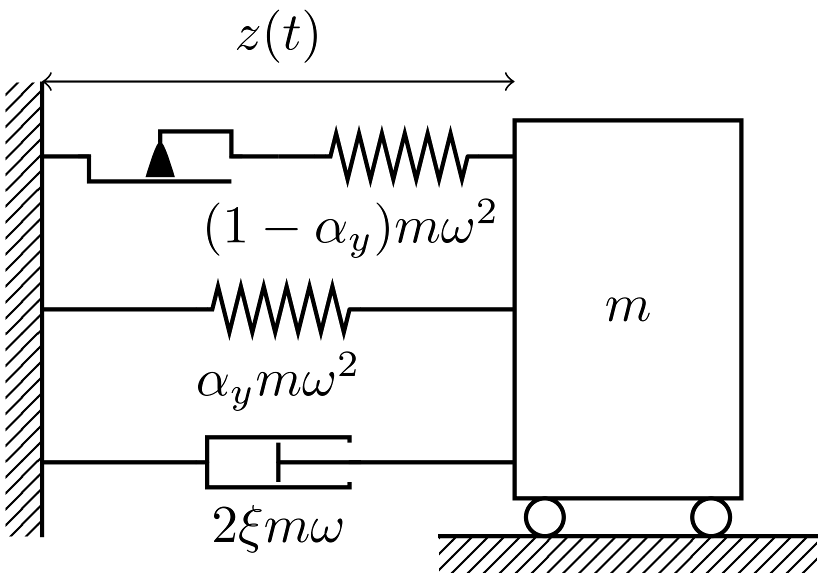
The first integral is approximated using Monte-Carlo simulation with a sample $(\boldsymbol{\alpha}_i, \boldsymbol{U}_i)_{1 \le i \le N}$ drawn from the product measure $\boldsymbol{h} \otimes \mathbb{P}_{\mathrm{X}}$.

The two expectations are approximated using Gauss-Hermite quadrature:

$$\mathbb{E}_{Z \sim \mathcal{N}(\mu, \sigma^2)}[f(Z)] \approx \frac{1}{\sqrt{\pi}} \sum_{q=1}^{Q} \omega_q f(z_q) \,,$$

$$z_q = \mu + \sqrt{2}\sigma u_q \,,$$

Numerical application



Nonlinear oscillator with kinematic hardening

$$\ddot{\boldsymbol{z}}(\boldsymbol{t}) + 2\boldsymbol{\xi}\boldsymbol{\omega}\dot{\boldsymbol{z}}(\boldsymbol{t}) + \boldsymbol{f}^{\mathrm{NL}}(\boldsymbol{z}(\boldsymbol{t})) = -\boldsymbol{s}(\boldsymbol{t})\ , \tag{1}$$

where $\boldsymbol{f}^{\mathrm{NL}}$ is a nonlinear restoring force.

Uncertain mechanical parameters

Table: Probabilistic model of $\mathbf{X}$ for the nonlinear oscillator.

Variable	Name	Mean
$\boldsymbol{m}$ (kg)	Mass of the system	**300**
$\boldsymbol{k}$ (N/m)	Stiffness	$\mathbf{2.7\ 10^5}$
$\boldsymbol{\xi}$ (1)	Damping ratio	**0.015**
$\boldsymbol{z_d}$ (m)	Yield displacement	$\mathbf{5\ 10^{-3}}$
$\boldsymbol{\alpha_y}$ (1)	Post-yield stiffness	$\mathbf{2\ 10^{-4}}$

The marginal distributions are uniforms with 15% coefficient of variation. The parameters are considered independent.

Benchmark settings

The input variables of the GP is composed of the subset $(\boldsymbol{PGA}, \boldsymbol{k}, \boldsymbol{m})$ (these are the most influential variables). $\mathbf{10^5}$ artificial ground motions and random draws of $\boldsymbol{k}$ and $\boldsymbol{m}$ are generated and the nonlinear oscillator is evaluated for each realization.

[3] M. Gu, J. Palomo, and J. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *The R Journal*, 11, 01 2018.
doi: 10.32614/RJ-2019-011

Benchmark settings

The input variables of the GP is composed of the subset $(\boldsymbol{PGA}, \boldsymbol{k}, \boldsymbol{m})$ (these are the most influential variables). $\mathbf{10^5}$ artificial ground motions and random draws of $\boldsymbol{k}$ and $\boldsymbol{m}$ are generated and the nonlinear oscillator is evaluated for each realization.

Goal: Estimation of the seismic fragility curve with failure threshold $\boldsymbol{C} = \mathbf{2.1}\boldsymbol{m}_{\boldsymbol{z_d}}$ ($\boldsymbol{m}_{\boldsymbol{z_d}}$ is the mean value of the yield displacement $\boldsymbol{z_d}$.)

[3] M. Gu, J. Palomo, and J. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *The R Journal*, 11, 01 2018.
doi: 10.32614/RJ-2019-011

Benchmark settings

The input variables of the GP is composed of the subset $(\boldsymbol{PGA}, \boldsymbol{k}, \boldsymbol{m})$ (these are the most influential variables). $\mathbf{10^5}$ artificial ground motions and random draws of $\boldsymbol{k}$ and $\boldsymbol{m}$ are generated and the nonlinear oscillator is evaluated for each realization.

Goal: Estimation of the seismic fragility curve with failure threshold $\boldsymbol{C} = \mathbf{2.1}\boldsymbol{m}_{\boldsymbol{z_d}}$ ($\boldsymbol{m}_{\boldsymbol{z_d}}$ is the mean value of the yield displacement $\boldsymbol{z_d}$.)

10 randomly chosen realizations are used for initialization.

[3] M. Gu, J. Palomo, and J. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *The R Journal*, 11, 01 2018.
doi: 10.32614/RJ-2019-011

Benchmark settings

The input variables of the GP is composed of the subset $(\boldsymbol{PGA}, \boldsymbol{k}, \boldsymbol{m})$ (these are the most influential variables). $\mathbf{10^5}$ artificial ground motions and random draws of $\boldsymbol{k}$ and $\boldsymbol{m}$ are generated and the nonlinear oscillator is evaluated for each realization.

Goal: Estimation of the seismic fragility curve with failure threshold $\boldsymbol{C} = \mathbf{2.1}\boldsymbol{m}_{z_d}$ ($\boldsymbol{m}_{z_d}$ is the mean value of the yield displacement $\boldsymbol{z}_d$.)

$\mathbf{10}$ randomly chosen realizations are used for initialization.

At step $\boldsymbol{n}$, $\boldsymbol{m} = \mathbf{1000}$ candidate points $(\boldsymbol{A}_i, \mathbf{X}_i)_{1\leq i\leq m}$ are subsampled in the dataset of $\mathbf{10^5}$ computations. We define:

$$(\boldsymbol{A}_{n+1}^{\mathrm{SUR}}, \mathbf{X}_{n+1}^{\mathrm{SUR}}) = \underset{1\leq i\leq m}{\operatorname{argmin}} \boldsymbol{J}_n(\boldsymbol{A}_i, \mathbf{X}_i)\ .$$

[3] M. Gu, J. Palomo, and J. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *The R Journal*, 11, 01 2018.
doi: 10.32614/RJ-2019-011

Benchmark settings

The input variables of the GP is composed of the subset $(\boldsymbol{PGA}, \boldsymbol{k}, \boldsymbol{m})$ (these are the most influential variables). $\mathbf{10^5}$ artificial ground motions and random draws of $\boldsymbol{k}$ and $\boldsymbol{m}$ are generated and the nonlinear oscillator is evaluated for each realization.

Goal: Estimation of the seismic fragility curve with failure threshold $\boldsymbol{C} = \mathbf{2.1}\boldsymbol{m}_{\boldsymbol{z_d}}$ ($\boldsymbol{m}_{\boldsymbol{z_d}}$ is the mean value of the yield displacement $\boldsymbol{z_d}$.)

$\mathbf{10}$ randomly chosen realizations are used for initialization.

At step $\boldsymbol{n}$, $\boldsymbol{m} = \mathbf{1000}$ candidate points $(\boldsymbol{A}_i, \mathbf{X}_i)_{1 \le i \le m}$ are subsampled in the dataset of $\mathbf{10^5}$ computations. We define:

$$(\boldsymbol{A}_{n+1}^{\mathrm{SUR}}, \mathbf{X}_{n+1}^{\mathrm{SUR}}) = \underset{1 \le i \le m}{\operatorname{argmin}} \, \boldsymbol{J}_n(\boldsymbol{A}_i, \mathbf{X}_i) \, .$$

The Gaussian process hyperparameters are updated every $\mathbf{10}$ iterations using a MAP estimator with a jointly robust prior [3].

[3] M. Gu, J. Palomo, and J. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *The R Journal*, 11, 01 2018.
doi: 10.32614/RJ-2019-011

Performance metrics

A numerical benchmark is carried out to compare the performance of SUR strategy and Monte-Carlo designs in terms of posterior variance:

$$\boldsymbol{v_n} = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{u}; \boldsymbol{G}) - \widehat{\boldsymbol{\Psi}}_n(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha})\boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u}) \middle| \mathcal{F}_n\right] ,$$

Performance metrics

A numerical benchmark is carried out to compare the performance of SUR strategy and Monte-Carlo designs in terms of posterior variance:

$$\boldsymbol{v_n} = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\boldsymbol{\mathcal{A}}\times\boldsymbol{\mathcal{X}}} (\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{u}; \boldsymbol{G}) - \widehat{\boldsymbol{\Psi}}_{\boldsymbol{n}}(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha}) \boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u}) \middle| \boldsymbol{\mathcal{F}_n}\right] ,$$

and in terms of bias using a reference fragility curve $\boldsymbol{\Psi}_{\mathrm{ref}}$:

$$\boldsymbol{b_n} = \int_{\boldsymbol{\mathcal{A}}\times\boldsymbol{\mathcal{X}}} (\boldsymbol{\Psi}_{\mathrm{ref}}(\boldsymbol{\alpha}, \boldsymbol{u}) - \widehat{\boldsymbol{\Psi}}_{\boldsymbol{n}}(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha}) \boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u}) ,$$

A numerical benchmark is carried out to compare the performance of SUR strategy and Monte-Carlo designs in terms of posterior variance:

$$\boldsymbol{v_n} = \mathbb{E}_{\mathbb{P}_0}\left[\int_{\mathcal{A}\times\mathcal{X}} (\boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{u}; \boldsymbol{G}) - \widehat{\boldsymbol{\Psi}}_n(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha})\boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u}) \middle| \mathcal{F}_n\right] ,$$

and in terms of bias using a reference fragility curve $\boldsymbol{\Psi}_{\mathrm{ref}}$:

$$\boldsymbol{b_n} = \int_{\mathcal{A}\times\mathcal{X}} (\boldsymbol{\Psi}_{\mathrm{ref}}(\boldsymbol{\alpha}, \boldsymbol{u}) - \widehat{\boldsymbol{\Psi}}_n(\boldsymbol{\alpha}, \boldsymbol{u}))^2 \boldsymbol{dh}(\boldsymbol{\alpha})\boldsymbol{d}\mathbb{P}_{\mathrm{X}}(\boldsymbol{u}) ,$$

The integral is evaluated with a Monte-Carlo sample of size 5000 and the expectation on $\mathbb{P}_0$ using 4000 realizations of the GP surrogate.

Benchmark setting

The SUR strategy is compared to a Monte-Carlo design.

Benchmark setting

The SUR strategy is compared to a Monte-Carlo design.

100 replications of Monte-Carlo designs for several training sizes are computed.

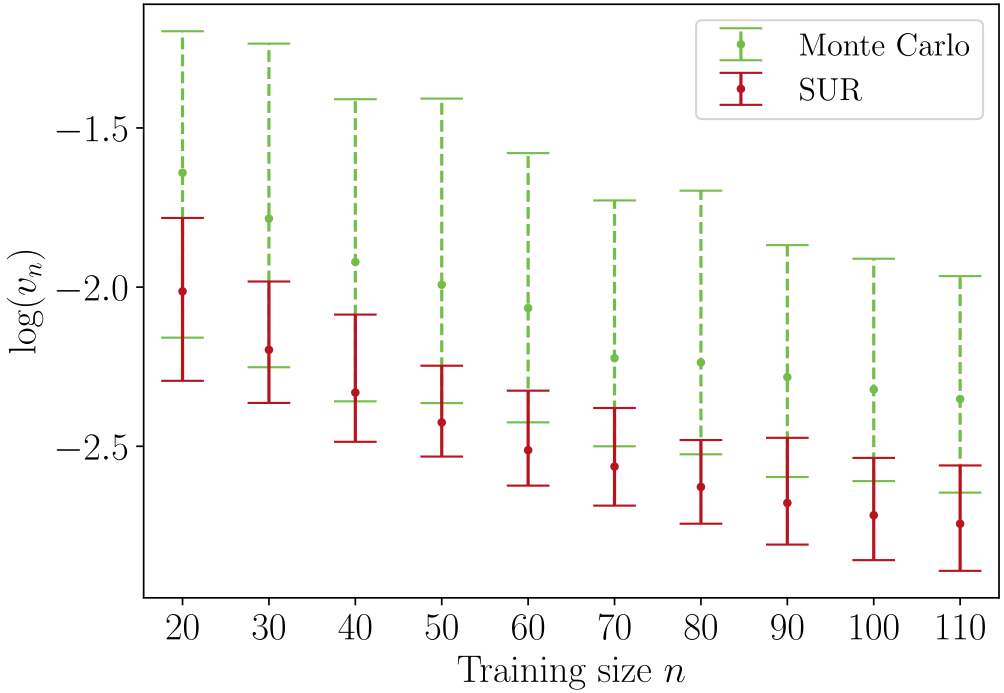
Benchmark setting

The SUR strategy is compared to a Monte-Carlo design.

100 replications of Monte-Carlo designs for several training sizes are computed.

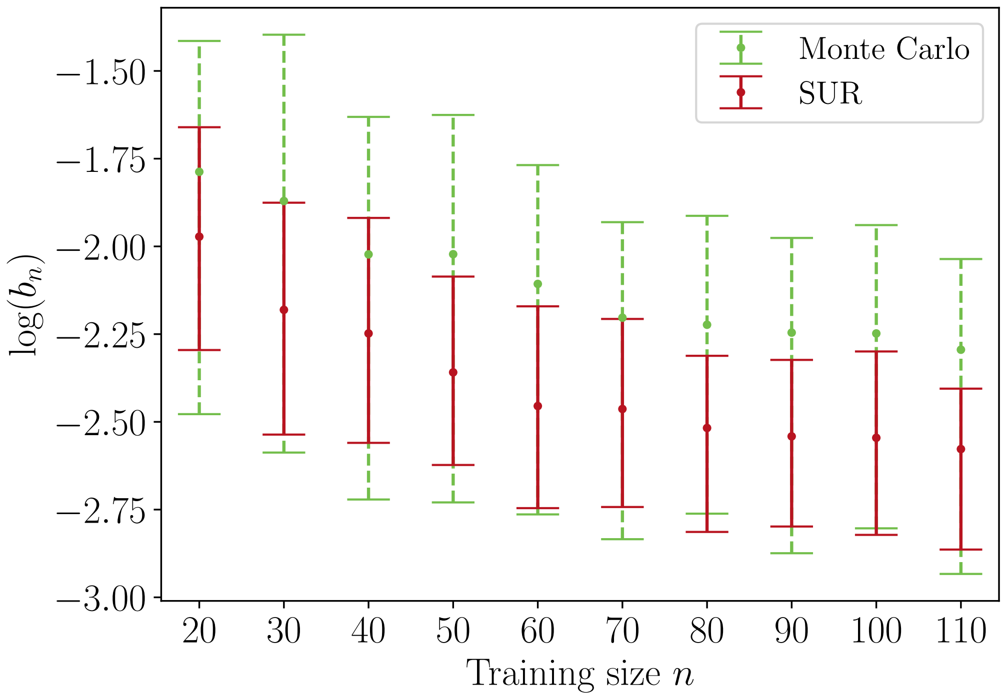
Due to the randomness induced in the SUR algorithm by choosing the candidate points at each step, **100** runs of the SUR strategy are carried out using HPC.

Performance assessment



Comparison of the posterior variance v_n between 100 Monte-Carlo designs and 100 runs of the SUR strategy for a failure threshold $C = 2.1 m_{z_d}$.

Performance assessment



Comparison of the posterior bias $\boldsymbol{b_n}$ between 100 Monte-Carlo designs and 100 runs of the SUR strategy for a failure threshold $\boldsymbol{C = 2.1 m_{z_d}}$.

Conclusion

SUR strategy is an heuristic to solve an intractable Bayesian decision problem.

Conclusion

SUR strategy is an heuristic to solve an intractable Bayesian decision problem.

Advantage: It defines a goal-oriented design of experiment strategy.

Conclusion

SUR strategy is an heuristic to solve an intractable Bayesian decision problem.

Advantage: It defines a goal-oriented design of experiment strategy.

Drawback: Very sensitive to the dimension of the input parameter, difficult optimization problem.

Merci pour votre attention !

`clgch.github.io`

Références

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. 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.

Clement Gauchy, Cyril Feau, and Josselin Garnier. Estimation of seismic fragility curves by sequential design of experiments. February 2022. hal-03588974.

M. Gu, J. Palomo, and J. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *The R Journal*, 11, 01 2018. doi: 10.32614/RJ-2019-011.

Backup