Engineering Risk Analysis Group
Department of Civil, Geo and Environmental Engineering
Technische Universität München

The cross-entropy method for model-based prediction and updating of rare events

RESIM 2021, 19 May 2021

Iason Papaioannou

Engineering Risk Analysis Group, TU München

E-mail: iason.papaioannou@tum.de

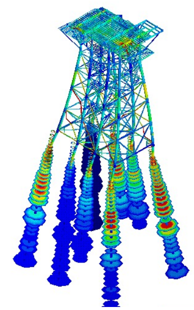
Joint Work with: Sebastian Geyer, Oindrila Kanjilal and Daniel Straub

Model-based prediction

Input

$\boldsymbol{X}$

Engineering Model $\mathcal{M}$



Response

$Y = \mathcal{M}(\boldsymbol{X})$

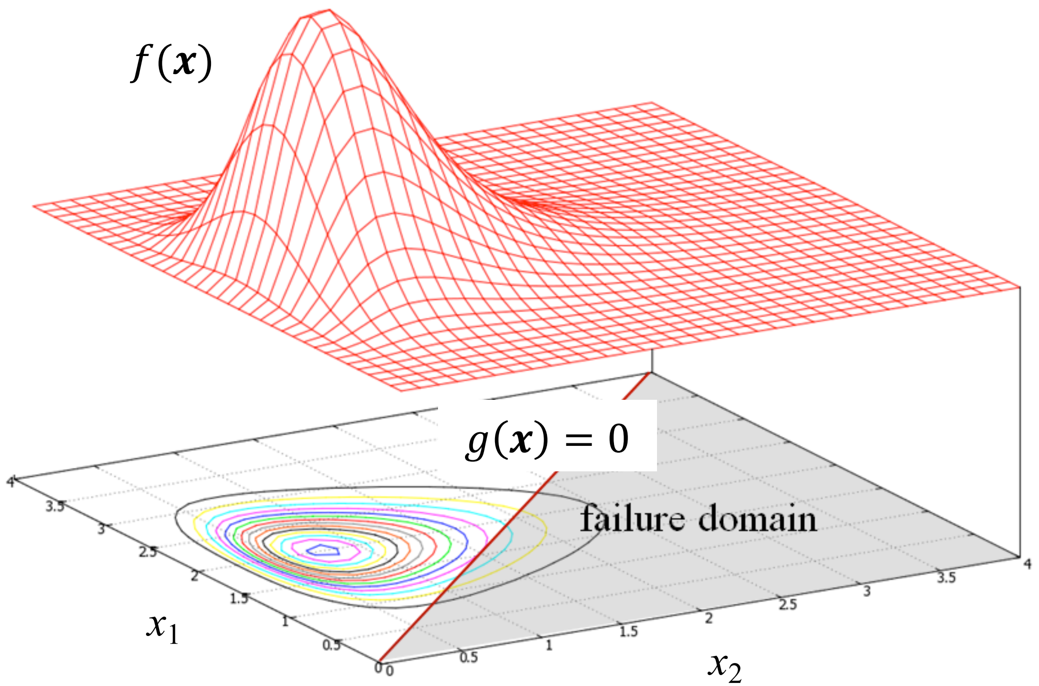
- $\mathcal{M}$: Model that often depends on the numerical solution of a PDE system
- $\boldsymbol{X} = [X_1; X_2; \dots; X_n]$: Vector of input random variables with joint PDF $f(\boldsymbol{x})$
- The model allows extrapolation to extreme situations where data is not available

Reliability analysis

Performance function $g(\boldsymbol{x}) = \tilde{g} \circ \mathcal{M}(\boldsymbol{x})$; Failure event $F = \{g(\boldsymbol{X}) \leq 0\}$

Probability of failure:

$$p_F := \mathbb{P}(F) = \int_{g(\boldsymbol{x})\leq 0} f(\boldsymbol{x}) d\boldsymbol{x} = \mathrm{E}_f[I(g(\boldsymbol{X}) \leq 0)]$$



Reliability methods

- Approximation methods based on Taylor series: FORM/SORM
- Sampling methods, e.g. splitting, subset simulation, cross-entropy method, line sampling, sequential importance sampling
- Surrogate-based reliability methods, e.g. Gaussian process (kriging), polynomial chaos expansion (PCE), neural networks, low-rank tensors

Importance sampling

Probability of failure

$$p_F = \mathrm{E}_h[I(g(\boldsymbol{X}) \le 0) w(\boldsymbol{X})]$$

Importance sampling density: $h(\boldsymbol{x})$

Importance weight function: $w(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{h(\boldsymbol{x})}$

Estimate of probability

$$\hat{p}_F = \widehat{\mathrm{E}}_h[I(g(\boldsymbol{X}) \le 0) w(\boldsymbol{X})] = \frac{1}{N} \sum_{k=1}^{N} I(g(\boldsymbol{x}_k) \le 0) w(\boldsymbol{x}_k), \qquad \boldsymbol{x}_k \sim h(\cdot)$$

Optimal (zero variance) IS density

$$h^*(\boldsymbol{x}) = \frac{1}{p_F} I(g(\boldsymbol{x}) \le 0) f(\boldsymbol{x})$$

Cross-entropy (CE) method [Rubinstein 1997]

Define a family of parametric densities $h(\boldsymbol{x};\boldsymbol{v})$, $\boldsymbol{v} \in \mathcal{V}$

Find parameters $\boldsymbol{v}$ by minimizing the Kullback-Leibler (K-L) divergence between $h^*(\boldsymbol{x})$ and $h(\boldsymbol{x};\boldsymbol{v})$:

$$\boldsymbol{v}^* = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmin}}\, D\big(h^*, h(\cdot;\boldsymbol{q})\big)$$

K-L divergence:

$$D\big(h^*, h(\cdot;\boldsymbol{q})\big) = \mathrm{E}_{h^*}\left[\ln\left(\frac{h^*(\boldsymbol{X})}{h(\boldsymbol{X};\boldsymbol{q})}\right)\right]$$

Substituting the density h^*, we get

$$\boldsymbol{v}^* = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmax}}\, \mathrm{E}_f\big[I(g(\boldsymbol{X}) \le 0)\ln\big(h(\boldsymbol{X};\boldsymbol{q})\big)\big]$$

Stochastic counterpart:

$$\widehat{\boldsymbol{v}} = \underset{\boldsymbol{q}\in\mathcal{V}}{\operatorname{argmax}} \sum_{k=1}^{N_{CE}} I(g(\boldsymbol{x}_k) \le 0)\ln\big(h(\boldsymbol{x}_k;\boldsymbol{q})\big)\ , \qquad \boldsymbol{x}_k \sim f(\cdot)$$

Multilevel CE method

Define a sequence of intermediate events

$$\{g(\boldsymbol{X}) \leq \gamma_t\}, t = 1, \ldots, T \quad \text{with} \quad \gamma_1 > \cdots > \gamma_T = 0$$

and a corresponding sequence of parameter vectors $\widehat{\boldsymbol{v}}_1, \ldots, \widehat{\boldsymbol{v}}_T$.

- Select γ_t such that for a not too small p^* (e.g., $p^* = 0.1$)

$$\mathbb{P}_{h(\cdot;\widehat{\boldsymbol{v}}_{t-1})}(g(\boldsymbol{X}) \leq \gamma_t) \geq p^*$$

- Solve the CE optimization program for estimating $\Pr(g(\boldsymbol{X}) \leq \gamma_t)$

$$\widehat{\boldsymbol{v}}_t = \underset{v \in \mathcal{V}}{\operatorname{argmax}} \sum_{k=1}^{N_{CE}} I(g(\boldsymbol{x}_k) \leq \gamma_t) \ln\big(h(\boldsymbol{x}_k; \boldsymbol{q})\big) w_t(\boldsymbol{x}_k; \widehat{\boldsymbol{v}}_{t-1}), \qquad \boldsymbol{x}_k \sim h(\cdot; \widehat{\boldsymbol{v}}_{t-1})$$

where $w_t(\boldsymbol{x}_k; \widehat{\boldsymbol{v}}_{t-1}) = \frac{f(\boldsymbol{x}_k)}{h(\boldsymbol{x}_k; \widehat{\boldsymbol{v}}_{t-1})}$

Improved CE (iCE) method [Papaioannou et al. 2019]

The multilevel CE method solves the CE optimization problem for a series of target densities $\{h_t, t = 1, \ldots, T\}$, with

$$h_t(\boldsymbol{x}) \propto I(g(\boldsymbol{x}) \leq \gamma_t) f(\boldsymbol{x}) \quad \text{with} \quad \gamma_1 > \cdots > \gamma_T = 0$$

Idea: Employ an alternative sequence of densities that makes better use of the intermediate samples

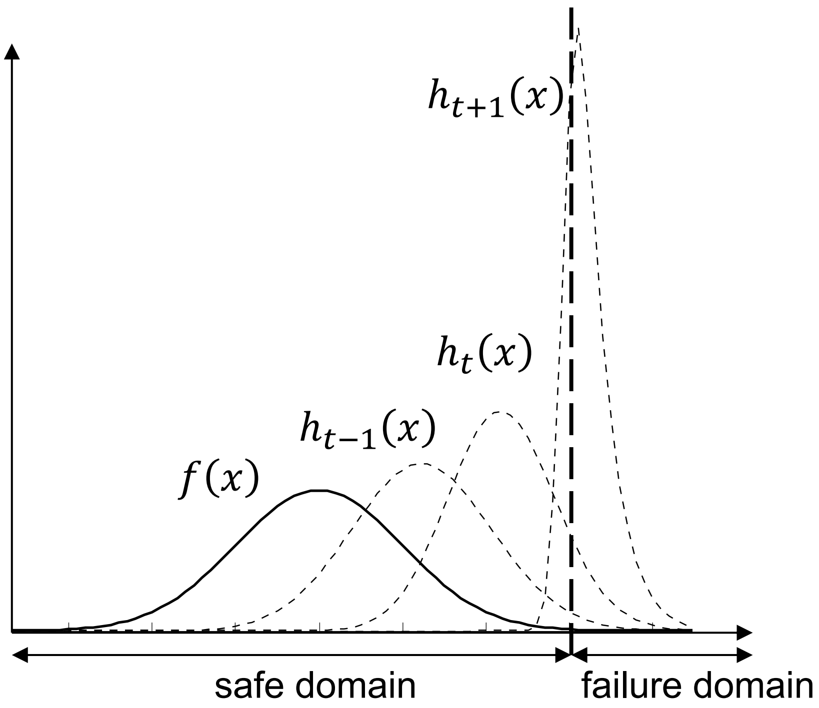
Define the intermediate densities s.t.

$$h_t(\boldsymbol{x}) \propto \Phi\left(-\frac{g(\boldsymbol{x})}{\sigma_t}\right) f(\boldsymbol{x}) \quad \text{with} \quad \sigma_1 > \cdots > \sigma_T > 0$$

where $\Phi(\cdot)$ is the standard normal CDF

iCE density sequence

$$h_t(\boldsymbol{x}) \propto \Phi\left(-\frac{g(\boldsymbol{x})}{\sigma_t}\right) f(\boldsymbol{x}) \quad \text{with} \quad \sigma_1 > \cdots > \sigma_T > 0$$



iCE algorithm

- Select σ_t such that for a target δ^* (e.g., $\delta^* = 1.5$)

$$\sigma_t = \underset{\sigma \in (0, \sigma_{k-1})}{\operatorname{argmin}} \left(\hat{\delta}_{\widetilde{w}_t}(\sigma) \; - \delta^* \right)^2$$

- Solve the CE optimization program with target density $h_t(\boldsymbol{x}) \propto \Phi\left(-\frac{g(\boldsymbol{x})}{\sigma_t}\right) f(\boldsymbol{x})$

$$\widehat{\boldsymbol{v}}_t = \underset{v \in \mathcal{V}}{\operatorname{argmax}} \sum_{k=1}^{N_{CE}} \ln\big(h(\boldsymbol{x}_k; \boldsymbol{q})\big) \widetilde{w}_t(\boldsymbol{x}_k; \widehat{\boldsymbol{v}}_{t-1}), \qquad \boldsymbol{x}_k \sim h(\cdot; \widehat{\boldsymbol{v}}_{t-1})$$

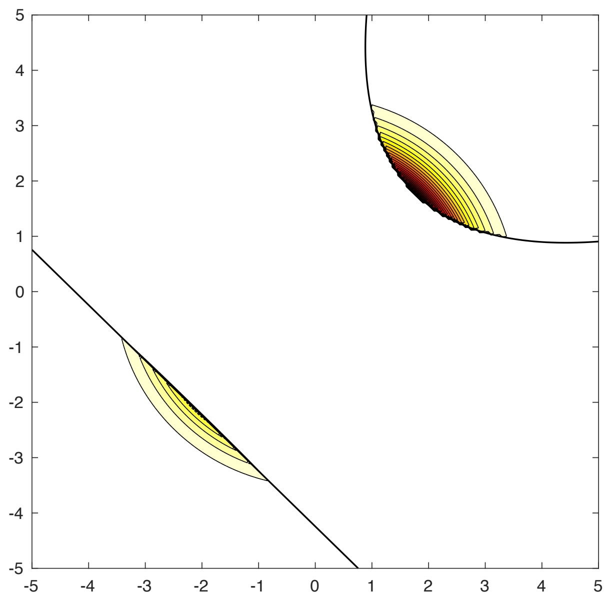
where $\widetilde{w}_t(\boldsymbol{x}_k; \widehat{\boldsymbol{v}}_{t-1}) = \frac{\Phi(-g(\boldsymbol{x}_k)/\sigma_t) f(\boldsymbol{x}_k)}{h(\boldsymbol{x}_k; \widehat{\boldsymbol{v}}_{t-1})}$

Choice of the parametric family

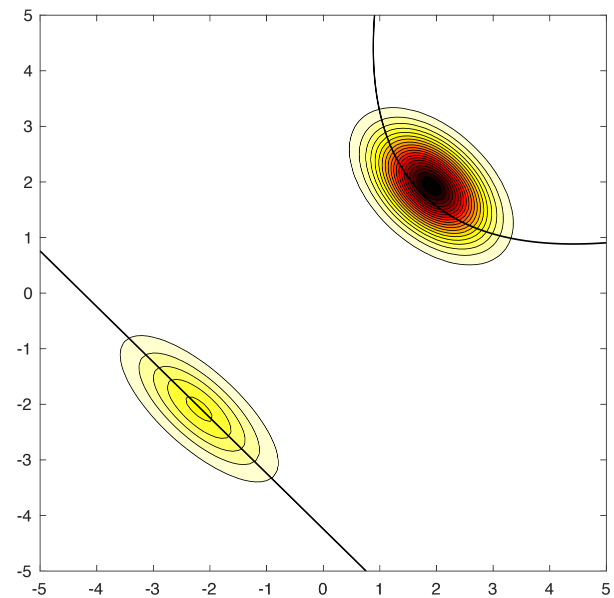
- Employ mixture models to treat multimodal failure domains
- Solution of the CE optimization problem with the EM algorithm [Geyer et a. 2019]
- Gaussian mixture model with # of parameters: $Kn(n+3)/2$

$$h(\boldsymbol{x};\boldsymbol{v}) = \sum_{i=1}^{K} \pi_i \varphi(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

Optimal density $h^*(\mathbf{x})$



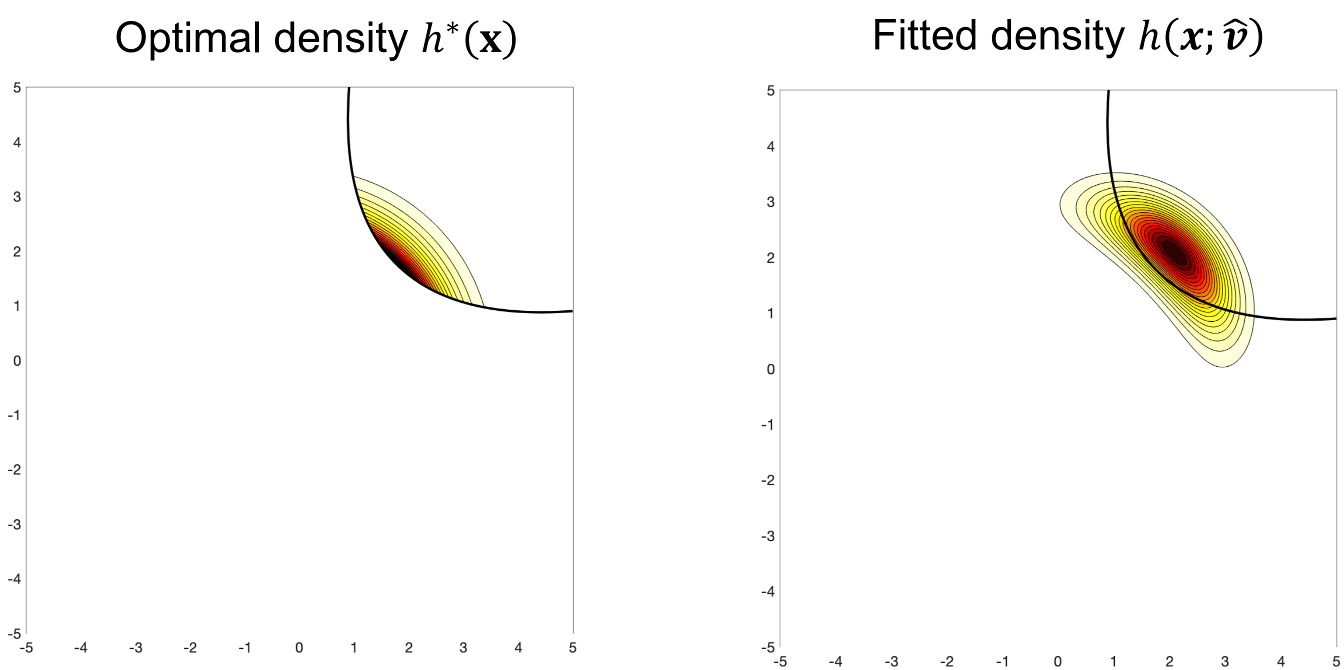
Fitted density $h(\boldsymbol{x};\hat{\boldsymbol{v}})$



A parametric family in high dimensions

- Assume that $\boldsymbol{X}$ is i.i.d. and express $\boldsymbol{X}$ in polar coordinates: $\boldsymbol{X} = R\boldsymbol{A}$
- Distance concentration phenomenon: $R/\mathrm{E}[R] \xrightarrow{\mathbb{P}} 1$, as $n \to \infty$
- von Mises-Fisher-Nakagami (vMFN) distribution model [Papaioannou et al. 2019]

$$h(r, \boldsymbol{a}; \boldsymbol{v}) = f_{\mathrm{N}}(r; [m, \Omega]) f_{\mathrm{vMF}}(\boldsymbol{a}; [\boldsymbol{\mu}, \kappa])$$



A parametric family in high dimensions

- Assume that $\boldsymbol{X}$ is i.i.d. and express $\boldsymbol{X}$ in polar coordinates: $\boldsymbol{X} = R\boldsymbol{A}$
- Distance concentration phenomenon: $R/\mathrm{E}[R] \xrightarrow{\mathbb{P}} 1$, as $n \to \infty$
- von Mises-Fisher-Nakagami (vMFN) distribution model [Papaioannou et al. 2019]
- vMFN mixture model with # of parameters: $K(n+3)$

$$h(r, \boldsymbol{a}; \boldsymbol{v}) = \sum_{i=1}^{K} \pi_i f_{\mathrm{N}}(r; [m_i, \Omega_{\mathrm{i}}]) f_{\mathrm{vMF}}(\boldsymbol{a}; [\boldsymbol{\mu}_i, \kappa_i])$$

Nonlinear SDOF oscillator

Equation of motion:

$$m\ddot{u}(t) + c\dot{u}(t) + k\left[\alpha u(t) + (1-\alpha)u_y z(t)\right] = f(t)$$

Bouc-Wen hysteresis law:

$$\dot{z}(t) = \frac{1}{u_y}\left[A\dot{u}(t) - \beta|\dot{u}(t)||z(t)|^{\tilde{n}-1}z(t) - \gamma\dot{u}(t)|z(t)|^{\tilde{n}}\right]$$

Discretized white noise load process:

$$f(t) = -m\sigma\sum_{i=1}^{n/2}\left[X_i\cos(\omega_i t) + X_{n/2+i}\sin(\omega_i t)\right]$$

with $\omega_i = i\Delta\omega; \Delta\omega = 30\pi/n$ (cut-off frequency $\omega_{\text{cut}} = 15\pi$); $\sigma = \sqrt{2S\Delta\omega}$; $S = 0.005\,\text{m}^2/\text{s}^3$; $\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$

Performance function: $g(\boldsymbol{x}) = u_y - u(8s)$

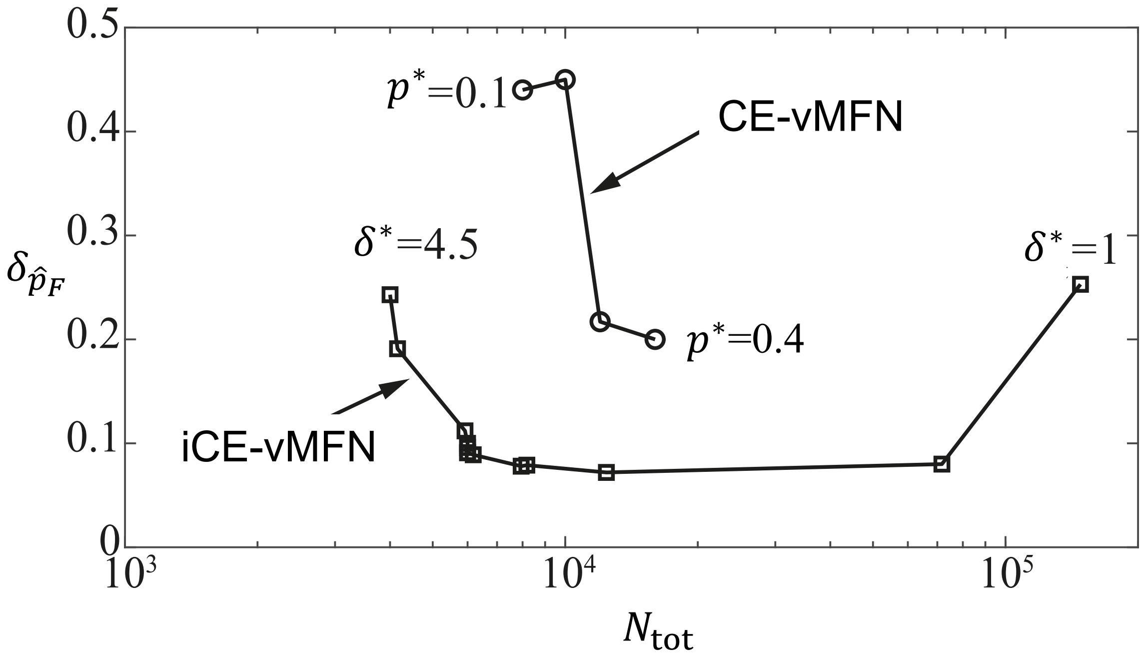
Nonlinear SDOF oscillator (II)

Number of random variables $n = 150$; vMFN distribution model

Number of samples per level $N_{CE} = 2000$ and $N = N_{CE}$

Reference probability $p_F = 6.28 \times 10^{-4}$

Vary p^* for standard CE and δ^* for iCE



Bayesian analysis

Assume that data $\boldsymbol{d}$ become available; $\boldsymbol{d}$ is described by the likelihood $L(\boldsymbol{x}|\boldsymbol{d})$

Posterior density of the random variables:

$$f_{\boldsymbol{d}}(\boldsymbol{x}) = \frac{1}{c_E} L(\boldsymbol{x}|\boldsymbol{d}) f(\boldsymbol{x})$$

Evidence (or marginal likelihood)

$$c_E = \mathrm{E}_f[L(\boldsymbol{X}|\boldsymbol{d})]$$

Updated (posterior) failure probability:

$$p_{F|\boldsymbol{d}} := \mathbb{P}(F|\boldsymbol{d}) = \mathrm{E}_{f_{\boldsymbol{d}}}[I(g(\boldsymbol{X}) \leq 0)] = \frac{\mathrm{E}_f[I(g(\boldsymbol{X}) \leq 0) L(\boldsymbol{X}|\boldsymbol{d})]}{\mathrm{E}_f[L(\boldsymbol{X}|\boldsymbol{d})]}$$

iCE for reliability updating [Kanjilal et al., in prep]

Estimation of denominator (evidence)

Optimal (zero variance) IS density

$$h_D^*(\boldsymbol{x}) = f_{\boldsymbol{d}}(\boldsymbol{x}) = \frac{1}{c_E} L(\boldsymbol{x}|\boldsymbol{d}) f(\boldsymbol{x})$$

Apply iCE with intermediate target densities defined as

$$h_{D,t}(\boldsymbol{x}) \propto L(\boldsymbol{x}|\boldsymbol{d})^{\beta_t} f(\boldsymbol{x}) \quad \text{with} \quad \beta_1 < \cdots < \beta_{T_D} = 1$$

Estimation of numerator

Optimal (zero variance) IS density

$$h_N^*(\boldsymbol{x}) \propto I(g(\boldsymbol{x}) \leq 0) L(\boldsymbol{x}|\boldsymbol{d}) f(\boldsymbol{x})$$

Apply iCE with intermediate target densities defined as

$$h_{N,t}(\boldsymbol{x}) \propto \Phi\left(-\frac{g(\boldsymbol{x})}{\sigma_t}\right) L(\boldsymbol{x}|\boldsymbol{d}) f(\boldsymbol{x}) \quad \text{with} \quad \sigma_1 > \cdots > \sigma_{T_N} > 0$$

Fatigue crack growth

Rate of crack growth of an infinite plate – Paris' law:

$$\frac{\mathrm{d}a(n)}{\mathrm{d}n} = C\left[\Delta S\sqrt{\pi a(n)}\right]^m$$

which can be solved to give:

$$a(n) = \left[\left(1-\frac{m}{2}\right)C\Delta S^m \pi^{\frac{m}{2}} n + a_0^{\left(1-\frac{m}{2}\right)}\right]^{\frac{1}{1-\frac{m}{2}}}$$

The number of stress cycles to reach a critical length a_c:

$$n_c = \frac{2}{(m-2)C(\sqrt{\pi}\Delta S)^m}\left[\frac{1}{a_0^{\frac{m-2}{2}}} - \frac{1}{a_c^{\frac{m-2}{2}}}\right], \quad m \neq 2$$

Performance function: $g = n_c - n_f$; with $n_f = 5\times10^5$ and $a_c = 50\mathrm{mm}$

Fatigue crack growth (II)

Prior distribution of model parameters $\boldsymbol{X} = [a_0; \Delta S; \ln C\,; m]$

Variable	Distribution	Mean	St. Dev.	Correlation
a_0 [mm]	Exponential	1	1	
ΔS [Nmm^{-2}]	Normal	60	10	
$\ln C\,, m[-]$	Bi-normal	$[-33; 3.5]$	$[0.47; 0.3]$	-0.9

The crack size is measured at $n_1 = 10^5$ and $n_2 = 2\times10^5$ as $a_{\mathrm{m},1} = 0.8\mathrm{mm}$, $a_{\mathrm{m},2} = 1.1\mathrm{mm}$

Likelihood function:

$$L(\boldsymbol{x}|\boldsymbol{d}) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{2}\left(\frac{a(\boldsymbol{x}, n_i) - a_{\mathrm{m}.i}}{\sigma_{\mathrm{m}}}\right)^2\right)$$

with $\sigma_{\mathrm{m}} = 0.2\mathrm{mm}$

Fatigue crack growth (III)

Results for $N = 500$, $\delta^* = 1.5$; Gaussian distribution model

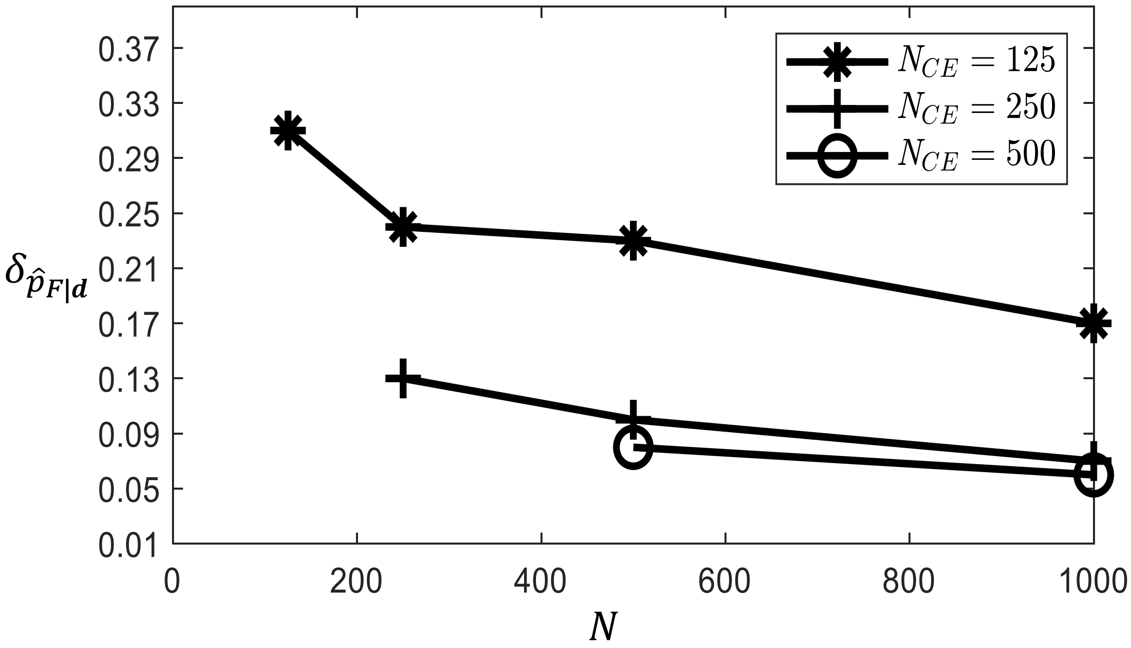
Reference probability: $p_{F|\boldsymbol{d}} = 4.9 \times 10^{-4}$

	N_{CE}		
	125	250	500
$\hat{p}_{F\|\boldsymbol{d}}$ $(\times 10^{-4})$	4.70	4.80	4.85
$\delta_{\hat{p}_{F\|\boldsymbol{d}}}$	0.23	0.10	0.08
T_D	2	2	2
T_N	7.3	6.3	5.9
N_{tot}	2038	2837	4443

Fatigue crack growth (III)

Results for $\delta^* = 1.5$; Gaussian distribution model

Reference probability: $p_{F|\boldsymbol{d}} = 4.9{\times}10^{-4}$



Summary

- iCE method employs a sequence of smooth approximations of the optimal IS
- The modified sequence enables efficient use of the intermediate samples to solve the CE optimization problems
- vMFN mixture distribution model extends the application of the CE method to high dimensional problems
- Application of the iCE method to Bayesian analysis of rare events

References

CE and iCE method

- Geyer, S., Papaioannou, I., & Straub, D. (2019). Cross entropy-based importance sampling using Gaussian densities revisited. *Structural Safety*, *76*, 15-27.
- Papaioannou, I., Geyer, S., & Straub, D. (2019). Improved cross entropy-based importance sampling with a flexible mixture model. *Reliability Engineering & System Safety*, *191*, 106564.
- Kanjilal, O., Papaioannou, I., & Straub, D. Structural reliability updating by cross entropy-based importance sampling. To appear in *Proc. 13th International Conference on Structural Safety and Reliability (ICOSSAR 2021)*.

Further developments of iCE

- Ehre, M., Papaioannou, I., Willcox, K. E., & Straub, D. (2021). Conditional reliability analysis in high dimensions based on controlled mixture importance sampling and information reuse. *Computer Methods in Applied Mechanics and Engineering*, *381*, 113826.
- Uribe, F., Papaioannou, I., Marzouk, Y. M., & Straub, D. (2021). Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation. To appear in *SIAM/ASA Journal on Uncertainty Quantification*. *arXiv preprint arXiv:2006.05496*.
- Kanjilal, O., Papaioannou, I., & Straub, D. (2021). Cross entropy-based importance sampling for first-passage probability estimation of randomly excited linear structures with parameter uncertainty. *Structural Safety*, *91*, 102090.