Uncertainty quantification (UQ) in geotechnical engineering

Ningxin Yang

Civil and Environmental Engineering, Imperial College London
n.yang23@imperial.ac.uk

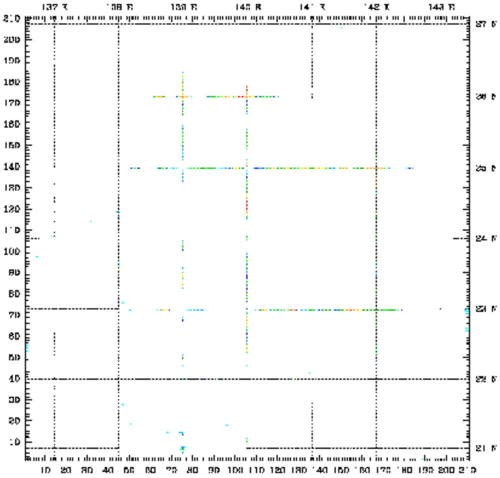
March 20, 2024

Supervisor: Dr Truong Le; Prof. Lidija Zdravkovic

Presentation Overview

Uncertainty sources

Real world with uncertainties



Source: Wikipedia

Types of two uncertainties

Aleatoric vs epistemic

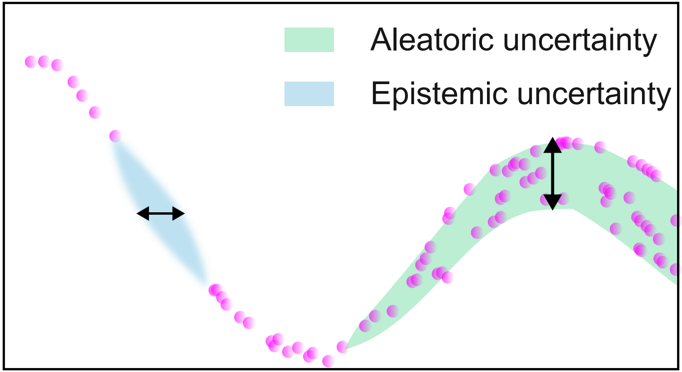


Figure: Aleatoric uncertainty vs Epistemic uncertainty

Distinguish:

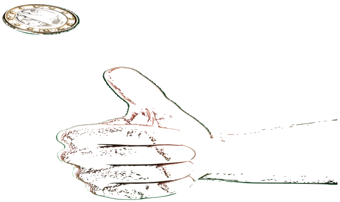
Aleatoric uncertainty

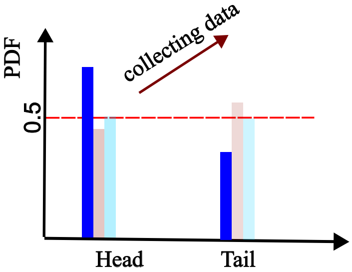
statistical variability, inherently random effects (**irreducible**)

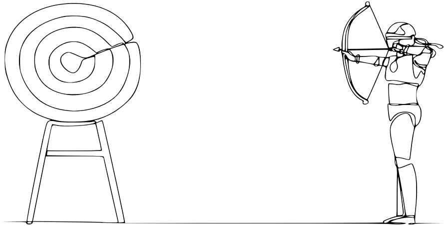
Epistemic uncertainty

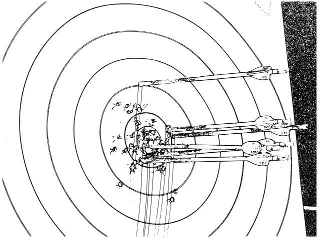
model uncertainty, a lack of knowledge (**reducible**)

Type one: aleatoric uncertainty

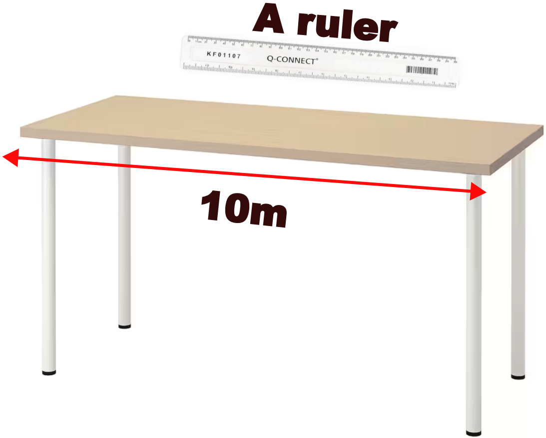








Type two: epistemic uncertainty



Random variables:

$x_i = x^\star(\text{10m}) \pm \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$

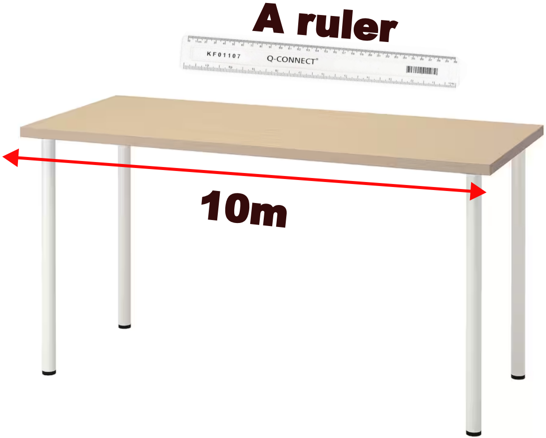
$x_i \sim iid$

$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$

First moment:

$$\begin{aligned} \mathbb{E}(\boldsymbol{x}) &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \mathbb{E}(x^\star)(\text{10m}) + \mathbb{E}(\boldsymbol{\epsilon}) \end{aligned} \tag{1}$$

Type two: epistemic uncertainty



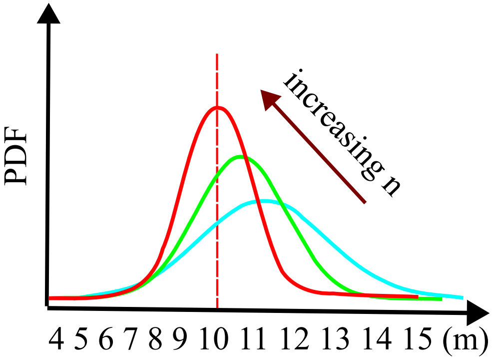
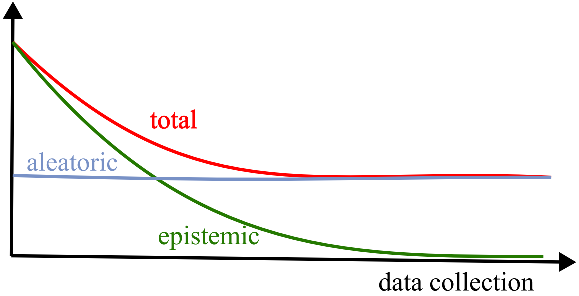


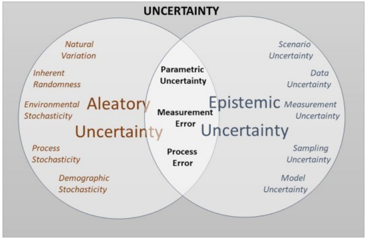
Figure: PDF with increasing trails

Uncertainty components:

Total uncertainty $\approx$ aleatoric uncertainty $+$ epistemic uncertainty



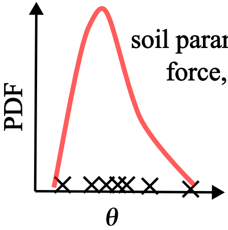
Not simple as it shows:

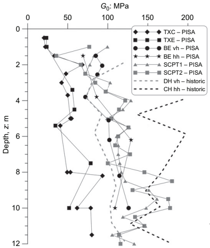


Source: *link*

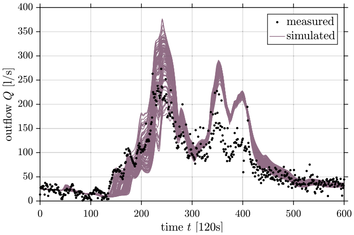
What uncertainties we have in geotechnical engineering exactly?

Mix of aleatoric and epistemic

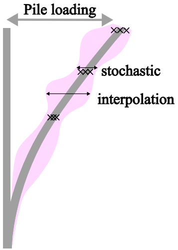




[1]



[2]

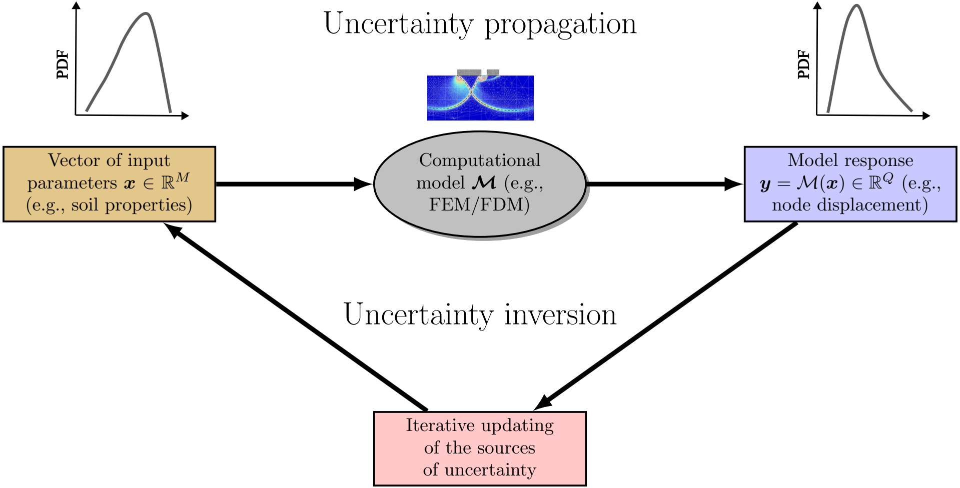


[1] Lidija Zdravković et al. “Ground characterisation for PISA pile testing and analysis”. In: *Géotechnique* 70.11 (2020), pp. 945–960.

[2] Joseph B Nagel, Jörg Rieckermann, and Bruno Sudret. “Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation”. In: *Reliability Engineering & System Safety* 195 (2020), p. 106737.

Some UQ problems and components

Two types of UQ problem

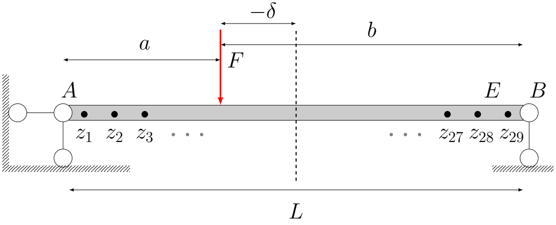


Component one: computational model

Definition

A computational model should contain:

- a mathematical description of the physics
- may be seen as a black box to compute the QoI



$$y = \begin{cases} \frac{Fbx[(L^2-b^2)-x^2]}{6LEI} & x \le a \\ \frac{Fb[\frac{L}{b}(x-a)^3+(L^2-b^2)]}{6LEI} & x > a \end{cases} \tag{1}$$

Component one: computational model

Definition

A computational model should contain:

- a mathematical description of the physics
- may be seen as a black box to compute the QoI

KJHH model in excavation in predict the deflection [3]:

$$\begin{aligned}\delta_{hm}(mm) =& a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + \\ & a_6X_1X_2 + a_7X_1X_3 + a_8X_1X_5\end{aligned} \quad (1)$$

where X_1 is the excavation depth, X_2 is $EI/\gamma_w h_{avg}^4$ support stiffness, ...

Component one: computational model

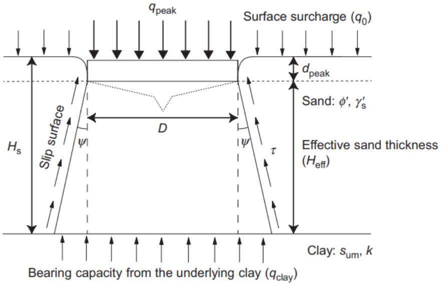
Definition

A computational model should contain:

- a mathematical description of the physics
- may be seen as a black box to compute the QoI

A model in predict peak resistance[4]:

$$
\begin{aligned}
q_{\text{peak, det}} = & (N_{c0}s_{\text{um}} + q_0 + 0{\cdot}12\gamma_s' H_s)\left(1 + \frac{1{\cdot}76H_s}{D}\tan\psi\right)^{E^*} \\
& + \frac{\gamma_s' D}{2\tan\psi(E^*+1)}\left[1 - \left(1 - \frac{1{\cdot}76H_s}{D}E^*\tan\psi\right) \right. \\
& \left. \times \left(1 + \frac{1{\cdot}76H_s}{D}\tan\psi\right)^{E^*}\right]
\end{aligned}
$$



Component one: computational model

Definition

A computational model should contain:

- a mathematical description of the physics
- may be seen as a black box to compute the QoI

Numerical methods:

- Finite element method
- Finite difference method
- . . .

Component two: probabilistic models of input parameters:

No data exist:

- expert judgment for selecting the input PDF's
- literature, data bases (e.g., on material properties)
- maximum entropy principle

Input data exist:

- classical statistical inference
- Bayesian statistics when data is scarce but there is some prior information

Data on output quantities:

- inverse probabilistic methods and Bayesian updating techniques

Component three: principles of uncertainty propagation:

Goal

estimate the uncertainty/variability of the *QoI* $\boldsymbol{Y} = \mathcal{M}(\boldsymbol{X})$ due to the uncertainty $f_{\boldsymbol{X}}$

- Output statistics, i.e., mean, standard deviation. etc.

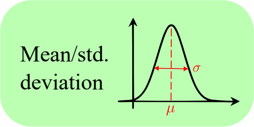
$$\mu_{\boldsymbol{Y}} = \mathbb{E}_{\boldsymbol{X}}[\mathcal{M}(\boldsymbol{X})]$$

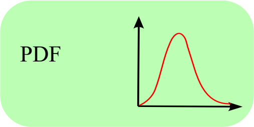
$$\sigma^2_{\boldsymbol{Y}} = \mathbb{E}_{\boldsymbol{X}}[(\mathcal{M}(\boldsymbol{X}) - \mu_{\boldsymbol{Y}})^2]$$

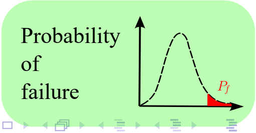
- Distribution of the QoI

- Probability of exceeding an admissible threshold y_{adm}

$$P_f = \mathbb{P}(\boldsymbol{Y} \geq y_{adm})$$







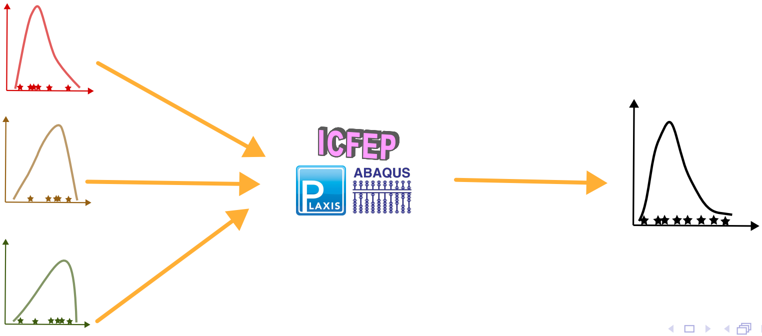
Uncertainty propagation based on Monte Carlo

Principle

Reproduce numerically the variability of the model parameters using a random number generator

Input space
(e.g., soil parameters)

Output space
(e.g., node displacements)



Pros and cons based on Monte Carlo:

PROS:

- Universal methods: only rely upon simulating random sampling and repeated computational model
- Suited to HPC
- Sound statistical foundations: convergence when $N_{MCS} \to \infty$

CONS:

- Statistical uncertainty: results are not exactly reproducible
- Low efficiency: convergence rate $\propto n^{-1/2}$

Example

An example To compute $P_f = 0.001$ is to be computed:

- At least 1,000 samples are needed in order to observe one single failure
- About 100 times more (i.e., 100,000 samples) are required to have a $\pm 10\%$ accuracy

Surrogate models for uncertainty quantification

Definition

A surrogate model $\tilde{\mathcal{M}}$ is an approximation of the original computational model $\mathcal{M}$ with the following features:

- It is built from a limited set of runs of the original model $\mathcal{M}$ called the experimental design $\mathcal{X} = \{\boldsymbol{x}^{(i)}, i = 1, \cdots, N\}$
- It assume some regularity of the model $\mathcal{M}$ and some general functional shape

Name	Shape	Parameters
Polynomial chaos expansions	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \boldsymbol{y}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$	$\boldsymbol{y}_{\boldsymbol{\alpha}}$
Low-rank tensor approximations	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{l=1}^{R} b_l \left(\prod_{i=1}^{M} v_l^i x_i \right)$	$b_l,\ z_{k,l}^i$
Kriging (a.k.a Gaussian processs)	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \boldsymbol{\beta}^T \cdot \boldsymbol{f}(\boldsymbol{x}) + Z(\boldsymbol{x}, \omega)$	$\boldsymbol{\beta},\ \sigma_Z^2,\ \boldsymbol{\theta}$
Support vector machines	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{i=1}^{m} a_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b$	$\boldsymbol{a},\ b$
Neural networks	$\tilde{\mathcal{M}}(\boldsymbol{x}) = f_n\left(\cdots f_2(b_2 + f_1(b_1 + \boldsymbol{w}_1 \cdot \boldsymbol{x}) \cdot \boldsymbol{w}_2)\right)$	$\boldsymbol{w},\ \boldsymbol{b}$

Advantages and challenges of a surrogate

Usage

$$\mathcal{M}(\boldsymbol{x}) \quad \approx \quad \tilde{\mathcal{M}}(\boldsymbol{x})$$

Hours to run — seconds to run

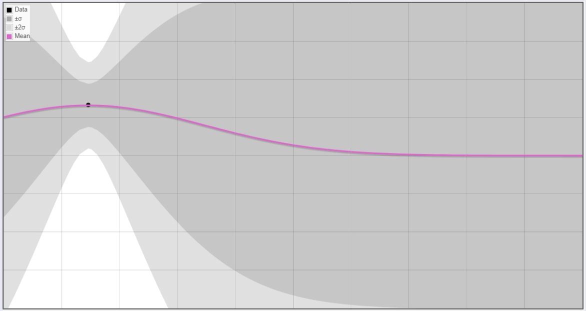
Advantages:

- Non-intrusive methods: Only based on the input-output of the computational model. such as MCS

Challenges:

- Need for a rigorous validation (overfitting or underfitting)
- Not free: Require too many EoD $\mathcal{X}$ if based on MCS-*curse of dimensionality*

Need for active learning



A flowchart for active learning

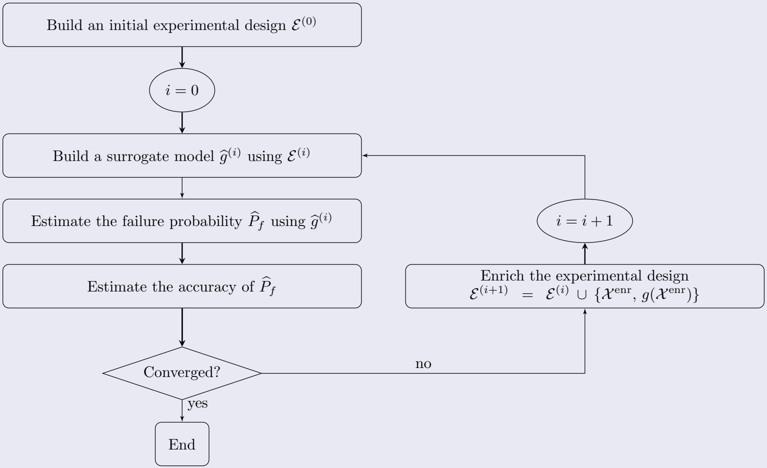


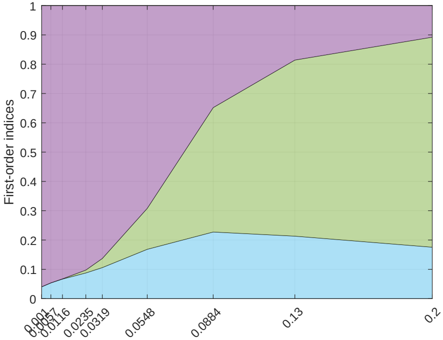
Figure: Active learning workflow [5]

Sensitivity analysis

Goal:

Determine what are the input parameters whose uncertainty explains the variability of the QoI

- detect input parameters whose uncertainty has no impact on the output variability
- detect input parameters which allow one to best decrease the output variability when set to a deterministic value
- detect interactions between paramters



Sobol's indice

Total variance:

$$D \equiv \text{Var}[\mathcal{M}(\boldsymbol{X})] = \text{Var}[\sum_{u \subset \{1,\cdots,M\}} \mathcal{M}_u(\boldsymbol{X}_u)] = \sum_{u \subset \{1,\cdots,M\}} \text{Var}[\mathcal{M}_u(\boldsymbol{X}_u)]$$

- Sobol's indice:

$$S_u \stackrel{\text{def}}{=} \frac{\text{Var}[\mathcal{M}_u(\boldsymbol{X}_u)]}{D}$$

- First-order Sobol's indice:

$$S_i = \frac{D_i}{D} = \frac{\text{Var}[\mathcal{M}_i(\boldsymbol{X}_i)]}{D}$$

  Quantify the effect of each input parameter separately
- Total Sobol's indice:

$$S_i^T \stackrel{\text{def}}{=} \sum_{u \supset i} S_u$$

  Quantify the total effect of x_i, including interactions with other variables

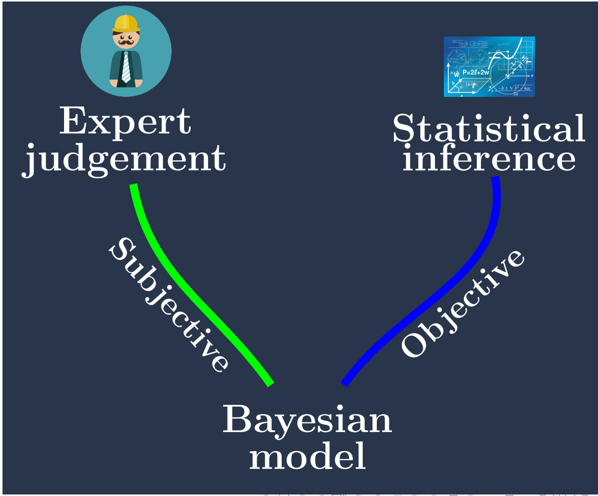
Inverse uncertainty

Choice for UQ inversion

Choice for the UQ method is totally based on the **quantity** of accessible data:

- **Lack** or **no** data available, model can be solely based on expert judgement
- **Substantial** volume data available, model can fully use statistical inference (e.g., the methods of moments)
- **Combination** of two above: Bayesian methods

$$\pi(\boldsymbol{x}|\mathcal{Y}) = \frac{\mathcal{L}(\boldsymbol{x}|\mathcal{Y}) \cdot \pi(\boldsymbol{x})}{\pi(\mathcal{Y})}$$



Posing the probabilistic problem

Uncertainties in math form:

$$\mathcal{Y}_i = \tilde{\mathcal{M}}(\boldsymbol{x}_i, \boldsymbol{\theta}^\star) + \delta(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i, i = 1, \cdots, N$$

$\mathcal{Y}_i$: observation; $\boldsymbol{x}_i$: control inputs; $\boldsymbol{\theta}^\star$: true value of calibrated parameters; $\tilde{\mathcal{M}}$: forward model; δ: model discrepancy; $\boldsymbol{\epsilon}_i$: observation error

Three speficic types of uncertainties (mixed with aleatoric and epistemic):

- observation error
- model discrepancy
- input space uncertainties

Should we consider aleatoric uncertainty (irreducible) in the inversion process?

- Or all burden will be shared by other epistemic uncertainty unreasonably

If only consider a Gaussian type observation error

$$\mathcal{Y}_i = \tilde{\mathcal{M}}(\boldsymbol{x}_i, \boldsymbol{\theta}^\star) + \boldsymbol{\epsilon}_i, i = 1, \cdots, N$$

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\mathcal{Y}) &= \prod_{i=1}^{N} N(\mathcal{Y}_i | \tilde{\mathcal{M}}(\boldsymbol{x}_i, \boldsymbol{\theta}), \boldsymbol{\Sigma}) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^{N_{\text{out}}} \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\left(\mathcal{Y}_i - \tilde{\mathcal{M}}(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left(\mathcal{Y}_i - \tilde{\mathcal{M}}(\boldsymbol{x}_i, \boldsymbol{\theta})\right)\right)
\end{aligned}$$

In real life, only one source of uncertainty is not convincing enough

$$\mathcal{Y}_i = \overset{\text{Model}}{\tilde{\mathcal{M}}(\boldsymbol{x}_i, \boldsymbol{\theta}^\star)} + \overset{\text{Model discrepancy}}{\delta(\boldsymbol{x}_i)} + \overset{\text{Observation error}}{\boldsymbol{\epsilon}_i}$$

$$\Sigma^{-1}$$

Bayesian inference

Difficulty with calculating evidence $\pi(\mathcal{Y})$

$$\pi(\boldsymbol{x}|\mathcal{Y}) = \frac{\mathcal{L}(\boldsymbol{x}|\mathcal{Y}) \cdot \pi(\boldsymbol{x})}{\pi(\mathcal{Y})} = \frac{\mathcal{L}(\boldsymbol{x}|\mathcal{Y}) \cdot \pi(\boldsymbol{x})}{\int_{\mathcal{D}_{\mathbf{X}}} \pi(\boldsymbol{x})\pi(\mathcal{Y}|\boldsymbol{x})\mathrm{d}\boldsymbol{x}}$$

Computing evidence $\pi(\mathcal{Y})$ is not a tractable problem. A common strategy is using *conjugate priors*

- static Bayesian network
- variant elimintion/ belief propagation
- kalman filtering

Computational methods:

$$\pi(\boldsymbol{x}|\mathcal{Y}) \approx \mathcal{L}(\boldsymbol{x}|\mathcal{Y}) \cdot \pi(\boldsymbol{x})$$

Samples from the posterior can be obtained through *Sampling methods* or *Optimization methods*

Definition

Optimisation based approximation: This method usually refers to variational inference. The basic principle is adopting some analytical distributions to approximate the posterior based on some loss functions

PROS and CONS:

- ✓ computational efficient and work well with large model
- ✓ has absolute converging criteria which makes easy t odetermine when to stop
- × unlikely to discover the globally optimal solution
- × precision constrained by the structure of approximaiton

Definition

Monte Carlo sampling: These techniques produce random samples from a proposal distribution, utilizing them to estimate both the posterior distribution and the associated statistics

Some are:

- probability density tranform
- rejection sampling
- importance sampling
- sequential Monte Carlo (e.g., particle filtering)
- Markov Chain Monte Carlo (MH, HMC, AIES)

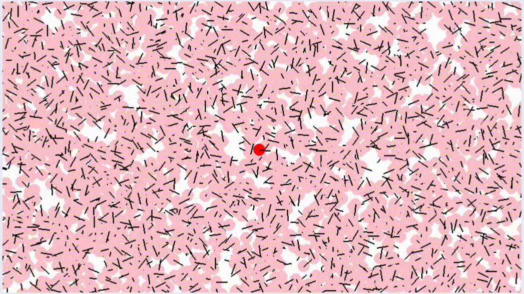
SMC vs MCMC

... non-iterative methods for generating independent samples..., we discuss an iterative method known as Markov Chain Monte Carlo, or MCMC for short, which produces dependent samples but which works well in high dimensions...[3]

[3] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

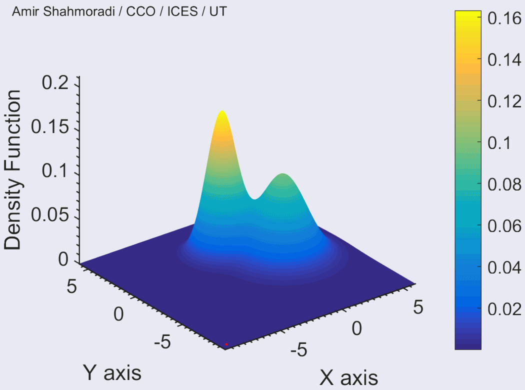
SMC vs MCMC

SMC



Source:weblink

MCMC



Source:weblink

Conclusion

Conclusion

1. In engineering, sometimes it is difficult to tell which uncertainty is aleatoric or epistemic or mixed type
2. Aleatoric uncertainty should be considered even it is irreducible
3. MCS is too expensive to force us to use a surrogate
4. Consider different types of uncertainty into Bayesian inference required linked with likelihood
5. MCMC is a good tool. But there exists some other alternatives.

References I

[1] Lidija Zdravković et al. "Ground characterisation for PISA pile testing and analysis". In: *Géotechnique* 70.11 (2020), pp. 945–960.

[2] Joseph B Nagel, Jörg Rieckermann, and Bruno Sudret. "Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation". In: *Reliability Engineering & System Safety* 195 (2020), p. 106737.

[3] C Hsein Juang et al. "Bayesian updating of soil parameters for braced excavations using field observations". In: *Journal of Geotechnical and Geoenvironmental Engineering* 139.3 (2013), pp. 395–406.

[4] Jinhui Li et al. "Bayesian prediction of peak resistance of a spudcan penetrating sand-over-clay". In: *Géotechnique* 68.10 (2018), pp. 905–917.

[5] Maliki Moustapha, Stefano Marelli, and Bruno Sudret. "Active learning for structural reliability: Survey, general framework and benchmark". In: *Structural Safety* 96 (2022), p. 102174.

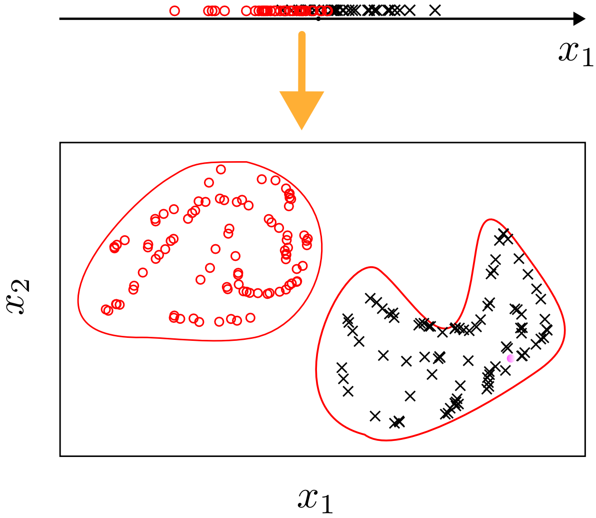
[6] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Thank you!

Questions? Comments?

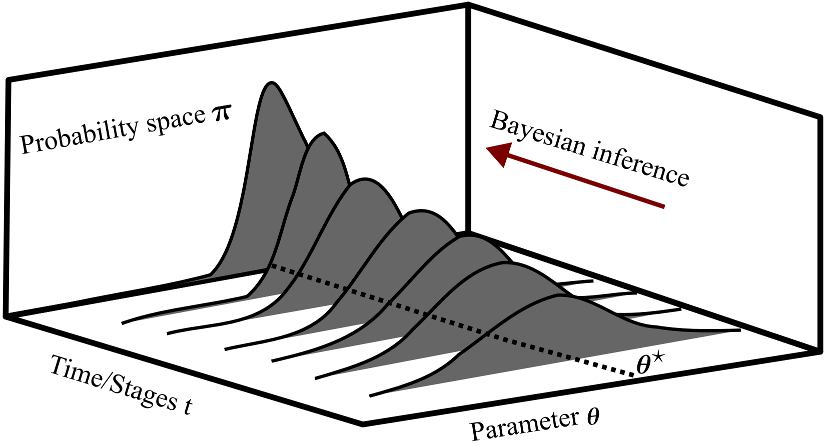
Appendix:

Aleatoric to epistemic

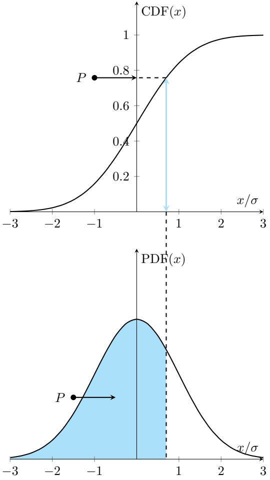


Appendix:

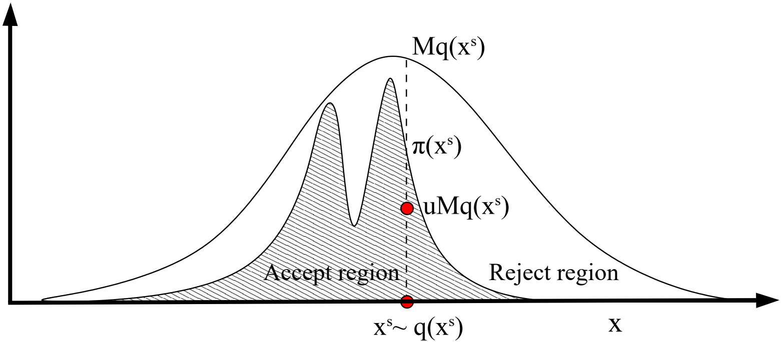
Sequential Bayesian inference



- Inverse probability transform



- Rejection sampling



$$E[f(\boldsymbol{x}_t)] = \int f(\boldsymbol{x}_t)\pi(\boldsymbol{x}_t)d\boldsymbol{x}_t \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{x}_t^s)$$

$$E[f(\boldsymbol{x}_t)] = \int f(\boldsymbol{x}_t)\pi(\boldsymbol{x}_t)d\boldsymbol{x}_t = \int f(\boldsymbol{x}_t)\frac{\pi(\boldsymbol{x}_t)}{q(\boldsymbol{x}_t)}q(\boldsymbol{x}_t)d\boldsymbol{x}_t \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{x}_t^s)\frac{\pi(\boldsymbol{x}_t^s)}{q(\boldsymbol{x}_t^s)}$$

- Require a proper proposal distribution $q(\boldsymbol{x}_t)$

$$f(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s) = \min(1, \alpha)$$

$$\alpha = \min(1, \frac{q(\boldsymbol{x}_t^s|\boldsymbol{x}_t^{s+1})\pi(\boldsymbol{x}_t^{s+1}|\mathcal{Y}_t)}{q(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s)\pi(\boldsymbol{x}_t^s|\mathcal{Y}_t)})$$

Appendix:
Metropolish-Hasting

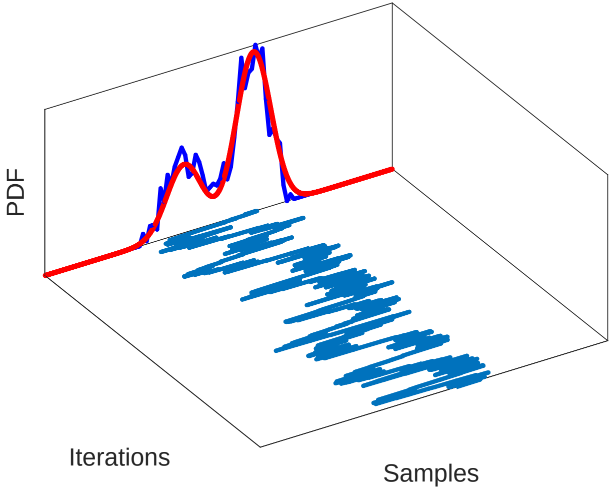
Algorithm 1: MH algorithm at t_{th} step

Data: $q(\boldsymbol{x}_t)$: Proposal distribution; $\pi(\boldsymbol{x}_t|\mathcal{Y})$: Target posterior.
Result: MCMC samples at t_{th} stage: $\mathcal{X}_t = \{\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^{N_\mathcal{X}}\}$

```
Initialization x_t^1 ∈ D_X;
for s ← 2 to N_X do
    Sample x_t^{s+1} ~ q(x_t^{s+1} | x_t^s);
    Compute acceptance probability α;
    Compute f(x_t^{s+1} | x_t^s) = min(1, α);
    Sample u ~ U(0, 1);
    Set candidate sample x_t^(⋆) to x_t^{s+1} with probability α;
end for
```

Appendix:

Metropolish-Hasting



$$\boldsymbol{x}_t^{(\star)} = \boldsymbol{x}_{t\_i}^{(s)} + z \cdot (\boldsymbol{x}_{t\_j}^{(\tilde{s})} - \boldsymbol{x}_{t\_i}^{(s)})$$

$$p(z|a) = \begin{cases} \frac{1}{\sqrt{z}(2\sqrt{a} - \frac{2}{\sqrt{a}})} & \text{if } z \in [1/a, a] \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha = \min(1, z^{M-1} \frac{\pi(x_t^{(\star)}|\mathcal{Y})}{\pi(x_{t\_i}^{(s)}|\mathcal{Y})})$$

Appendix:

AIES

Algorithm 2: AIES algorithm at t_{th} step

Data: $\pi(\boldsymbol{x}_t|\mathcal{Y})$: Target posterior; tuning parameter a

Result: MCMC samples at t_{th} stage: $\mathcal{X}_t = \{\mathcal{X}_{t\_1}, \cdots, \mathcal{X}_{t\_N_{chain}}\}$), with $\mathcal{X}_{t\_i} = \{\boldsymbol{x}^1_{t\_i}, \cdots, \boldsymbol{x}^{N_\mathcal{X}}_{t\_i}\}$

Initialization N_{chain} samples $\{\boldsymbol{x}^1_{t\_1}, \cdots, \boldsymbol{x}^1_{t\_N_{chain}}\}$, with $\boldsymbol{x}^1_{t\_i} \in \mathcal{D}_{\boldsymbol{X}}$

for $s \leftarrow 2$ *to* $N_\mathcal{X}$ **do**

  for $i \in \{1, \cdots, N_{chain}\}$ **do**

    Pick random j from $\{1, \cdots, N_{chain}\}\backslash i$;

    Propose $x_t^{(\star)}$ with **??**;

    Set $\boldsymbol{x}^s_{t\_i} = x_t^{(\star)}$ with probability α (see **??**);

  end for

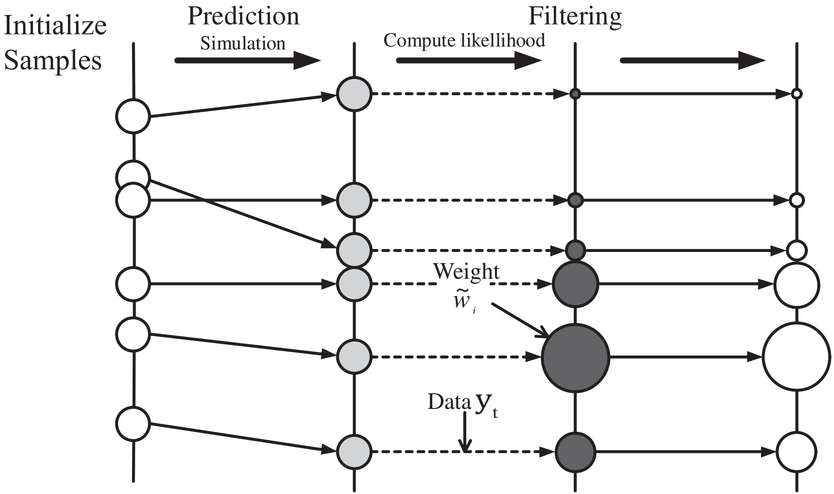
end for

$$\boldsymbol{x_t} = g(\boldsymbol{x_{t-1}}) + \boldsymbol{v} \qquad \text{(state equation)}$$
$$\mathcal{Y}_t = m(\boldsymbol{x_t}) + \boldsymbol{w} \qquad \text{(observation equation)}$$

$$\pi(\boldsymbol{x}_{1:t}|\mathcal{Y}_{1:t}) \approx \sum_{s=1}^{S} \tilde{w}_t^s \delta_{\boldsymbol{x}_{1:t}^s}(\boldsymbol{x}_{1:t})$$

$$\tilde{w}_t^s = \frac{w_t^s}{\sum_{s=1}^{S}(w_t^s)}$$

$$\pi(\boldsymbol{x}_{1:t}|\mathcal{Y}_{1:t}) \propto \pi(\mathcal{Y}_t|\boldsymbol{x}_t)\pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\pi(\boldsymbol{x}_{t-1}|\mathcal{Y}_{t-1})$$



Initialize
Samples
Prediction
Simulation
Filtering
Compute likellihood
Weight
$\tilde{w}_i$
Data $\mathbf{y}_t$

Algorithm 3: SISR algorithm at t_{th} step

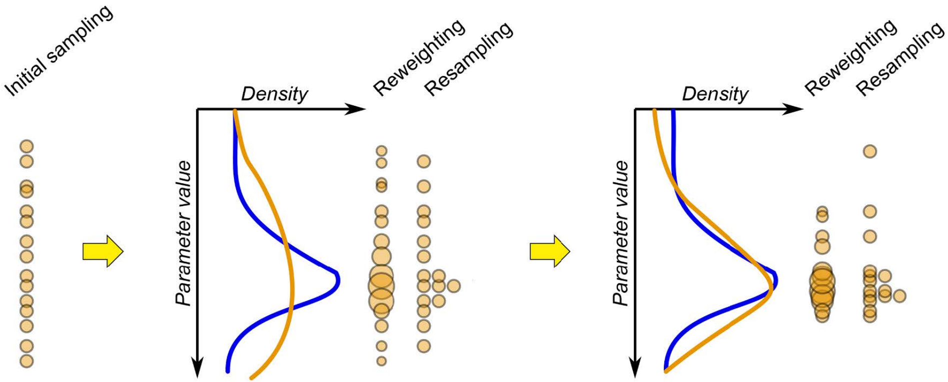
Data: Samples $\boldsymbol{x}_{t-1}^{s}$ with weights $w_{t-1}^{s},\ s = \{1, \cdots, N\}$; observation $\mathcal{Y}_t$ at t_{th} stage

Result: SMC samples with normalized weights $\tilde{w}_t^s$ at t_{th} stage: $\boldsymbol{x}_t^{(\star)} = \{\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^N\}$)

for $s \leftarrow 1$ *to* N **do**
  Sample from proposal distribution $\boldsymbol{x}_t^s \sim q(\boldsymbol{x}_t^s | \boldsymbol{x}_{t-1}^s, \mathcal{Y}_t)$;
  Compute weight using **??**;
end for
Normalized weights;
Calculate degeneracy measure using **??**;
if $\hat{S}_{eff} < S$ **then**
  Resample;
end if

Appendix:

Sequential Monte Carlo with resampling



Appendix:

Visualized PCE