Connections Between Rare-event Simulation and Bayesian inference

F.A. DiazDelaO

Institute for Risk and Uncertainty

About myself

- BSc in Applied Mathematics (ITAM)
- MSc in Statistics and Operational Research (Essex)
- PhD (Sheffield, 2007/Swansea)
- Postdoc (Swansea)
- Senior Lecturer. Institute for Risk and Uncertainty, School of Engineering (Liverpool, from 2018)
  - Collaborations with GE, Parker-Hannifin, Airbus, etc.
  - 3 PhD students, 1 Postdoc
  - EPSRC Fellow 2018-2021
  - Currently visiting fellow at the Alan Turing Institute.

Outline

Motivation

- Three important and challenging problems in modern science:
  1. Identifying model parameters
  2. Rating competitive models based on measured data
  3. Estimating the probability of failure of a system
- By solving them we can perform structural system identification, develop high fidelity models, design robust structures, amongst many other things.
- Two questions:
  1. Can a **link** be established between the Bayesian updating problem and the engineering reliability problem?
  2. If so, can we develop a robust and efficient **algorithm**?

The Bayesian Updating Problem

- Let $\mathcal{D}$ be an experimental dataset and θ be the parameters of model $\mathcal{M}$.
- Let $\mathcal{P}(\theta|\mathcal{M})$ be the prior distribution of θ.
- **Aim**: to find the posterior distribution of θ given $\mathcal{D}$ and $\mathcal{M}$.

$$\mathcal{P}(\theta|\mathcal{D},\mathcal{M}) = \frac{\mathcal{P}(\mathcal{D}|\theta,\mathcal{M})\mathcal{P}(\theta|\mathcal{M})}{\mathcal{P}(\mathcal{D}|\mathcal{M})}$$

with $\mathcal{P}(\mathcal{D}|\mathcal{M}) = \int \mathcal{P}(\mathcal{D}|\theta,\mathcal{M})\mathcal{P}(\theta|\mathcal{M})d\theta$.

- $\mathcal{P}(\mathcal{D}|\mathcal{M})$ is immaterial in this inference problem, but it is the main quantity of interest in model class selection as it provides the evidence.

The Engineering Reliability Problem

- Let $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}$ be a system performance function.
- **Aim**: To estimate the **probability of failure**, i.e. the probability of demand exceeding the capacity of the system.
- Let y^* be a critical value such that the system fails if $y = \mathcal{G}(x_1, \ldots, x_d) > y^*$.
- The **failure domain** F can thus be defined as:

$$F = \{x : \mathcal{G}(x) > y^*\}$$

- The engineering reliability problem can be formulated as computing the probability of failure:

$$p_F = \mathcal{P}(x \in F) = \int_F \pi(x) dx$$

Subset Simulation

- Developed by Au and Beck (2001) to simulate rare events and estimate small probabilities of failure.
- The idea is to decompose a rare event F into a sequence of progressively less rare events as:

$$F = F_m \subset F_{m-1} \subset \ldots \subset F_1$$

  where F_1 is a relatively frequent event.
- Given the above sequence of events, the small probability $\mathcal{P}(F)$ of the rare event can be represented as a product of larger probabilities as:

$$\mathcal{P}(F) = \mathcal{P}(F_m) = \mathcal{P}(F_1) \cdot \mathcal{P}(F_2|F_1) \cdot \ldots \cdot \mathcal{P}(F_m|F_{m-1})$$

Subset Simulation

- Subset simulation explores the input space $\mathcal{X}$ by generating a relatively small number of i.i.d. samples $x_0^{(1)}, \ldots, x_0^{(n)} \sim \pi(x)$ and computing the corresponding system responses $y_0^{(1)}, \ldots, y_0^{(n)}$.

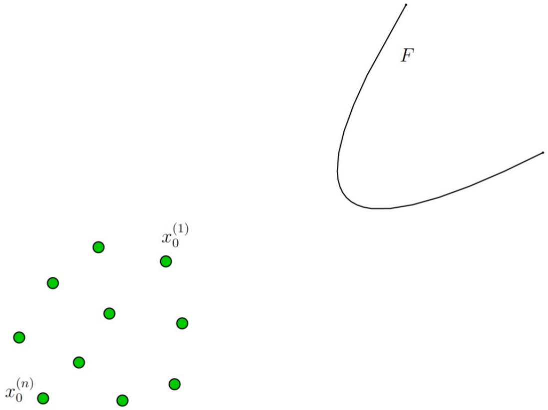


Figure taken from Zuev (2015).

Subset Simulation

- Let $p \in (0,1)$ such that $np \in \mathbb{N}$. Define the first intermediate failure domain as:

$$F_1 = \left\{ x : \mathcal{G}(x) > y_1^* = \frac{y_0^{(np)} + y_0^{(np+1)}}{2} \right\}$$

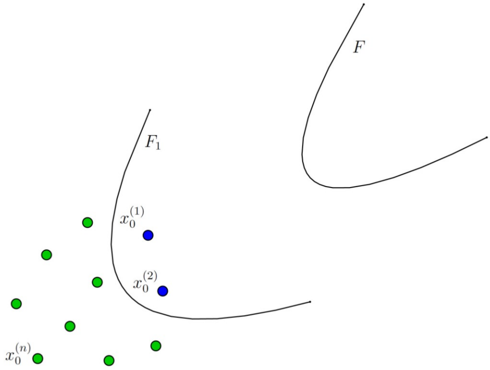


Figure taken from Zuev (2015).

Subset Simulation

- By construction, $x_0^{(1)}, \dots, x_0^{(np)} \in F_1$, whilst $x_0^{(np+1)}, \dots, x_0^{(n)} \notin F_1$.
- Thus, the Monte Carlo estimate for the probability of F_1 is given by

$$\mathcal{P}(F_1) \approx \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_{F_1}(x_0^{(i)}) = p$$

- F_1 provides a rough estimate to the failure domain F.
- Since $F \subset F_1$, the failure probability can be written as:

$$p_F = \mathcal{P}(F_1)\mathcal{P}(F|F_1)$$

- In the next stage, instead of sampling in the whole input space, SuS populates F_1.

Subset Simulation

- We start with $x_0^{(1)}, \ldots, x_0^{(np)} \sim \pi(x|F_1)$ and need to draw $n - np$ samples from $\pi(x|F_1)$.
- This is done with an MCMC scheme.

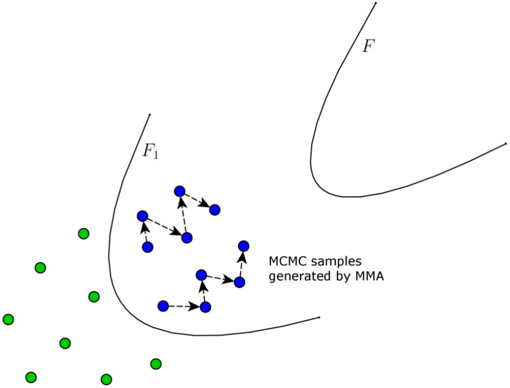


Figure taken from Zuev (2015).

Subset Simulation

- Define the second intermediate failure domain as:

$$F_2 = \left\{ x : \mathcal{G}(x) > y_2^* = \frac{y_1^{(np)} + y_1^{(np+1)}}{2} \right\}$$

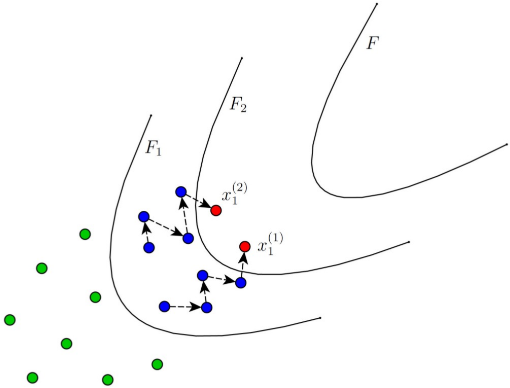


Figure taken from Zuev (2015).

Subset Simulation

- By construction, $x_1^{(1)}, \ldots, x_1^{(np)} \in F_2$, whilst $x_1^{(np+1)}, \ldots, x_1^{(n)} \notin F_2$.
- Thus, the Monte Carlo estimate for the probability of F_2 given F_1 is equal to

$$\mathcal{P}(F_2|F_1) \approx \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}_{F_2}(x_0^{(i)}) = p$$

- Since $F \subset F_2 \subset F_1$, the failure probability can be written as:

$$\begin{aligned} p_F &= \mathcal{P}(F_1)\mathcal{P}(F|F_1) \\ &= \mathcal{P}(F_1)\mathcal{P}(F_2|F_1)\mathcal{P}(F|F_2) \end{aligned}$$

- In the next stage, instead of sampling in the whole input space, SuS populates F_2.

Modified Metropolis Algorithm

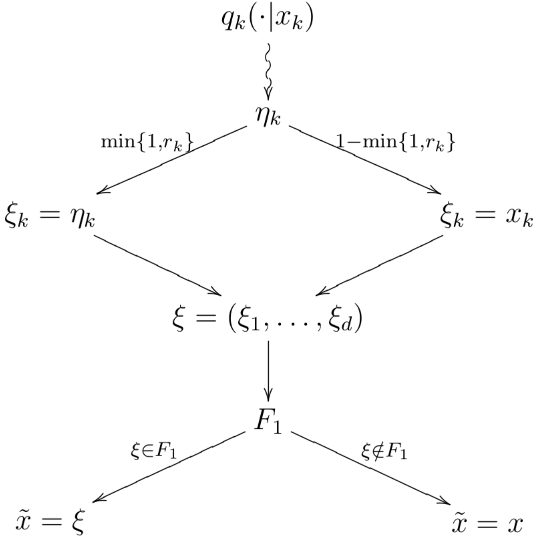


Figure taken from Zuev (2015).

Stopping Criterion

- The number of failure samples at the ℓ-th level is given by $n_F(\ell) = \sum_{i=1}^{n} \mathcal{I}_F(x_\ell^{(i)})$. Observe the following:
  - It is likely that $n_F(\ell) = 0$ for the first levels.
  - In general, $n_F(\ell) \geq n_F(\ell - 1)$.
  - $p_F = \mathcal{P}(F_1) \cdot \mathcal{P}(F_2|F_1) \cdot \ldots \cdot \mathcal{P}(F_\ell|F_{\ell-1}) \cdot \mathcal{P}(F|F_\ell)$
  - $p_F \approx p^\ell \mathcal{P}(F|F_\ell)$
- The last term is estimated as $\mathcal{P}(F|F_\ell) \approx \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_F(x_\ell^{(i)}) = \dfrac{n_F(\ell)}{n}$.
- If $\frac{n_F(\ell)}{n} \geq p$, then there are at least np failure samples. The current conditional level becomes the last level and the failure probability estimate becomes

$$p_F \approx p_F^{SuS} = p^\ell \frac{n_F(\ell)}{n}$$

- Otherwise, define the next intermediate failure domain $F_{\ell+1}$.

Bayesian Updating with Structural reliability methods

- BUS (Straub and Papaioannou, 2014) connects the Bayesian updating and structural reliability problems.

Algorithm 1 Rejection Sampling

Input: Prior distribution $\pi(\theta)$
Likelihood $\mathcal{L}(\theta) = \mathcal{P}(\mathcal{D}|\theta)$
$u \sim \mathcal{U}[0, 1]$
$c \in \mathbb{R}$ such that $c\mathcal{L}(\theta) < 1$

Output: Posterior distribution $\pi(\theta|\mathcal{D})$

1: Draw θ from $\pi(\theta)$, and u from $\mathcal{U}[0, 1]$
2: **if** $u < c\mathcal{L}(\theta)$ **then**
3: Accept θ
4: **else**
5: Go to Step 1
6: **end if**

Proof

- $\theta \sim \pi(\theta)$, $u \sim \mathcal{U}[0,1] = \mathcal{I}(0 \leq u \leq 1)$
- The joint pdf of θ and u is $\pi(\theta)\mathcal{I}(0 \leq u \leq 1)$
- The algorithm only accepts if $u < c\mathcal{L}(\theta)$, and produces

$$\begin{aligned} p(u,\theta) &= \frac{\pi(\theta)\mathcal{I}(0 \leq u \leq 1)\mathcal{I}(u < c\mathcal{L}(\theta))}{\int\int \pi(\theta)\mathcal{I}(0 \leq u \leq 1)\mathcal{I}(u < c\mathcal{L}(\theta))du d\theta} \\ &= \mathcal{P}_F^{-1}\pi(\theta)\mathcal{I}(0 \leq u \leq 1)\mathcal{I}(u < c\mathcal{L}(\theta)) \end{aligned}$$

- Thus, the marginal of θ is

$$\begin{aligned} p(\theta) &= \int_0^1 p(u,\theta)du \\ &= \mathcal{P}_F^{-1}\pi(\theta)\int_0^1 \mathcal{I}(0 \leq u \leq 1)\mathcal{I}(u < c\mathcal{L}(\theta))du \\ &= \mathcal{P}_F^{-1}\pi(\theta)c\mathcal{L}(\theta) \propto \pi(\theta|\mathcal{D}) \end{aligned}$$

BUS

- The acceptance rate is generally low $\Rightarrow$ use subset simulation!
- The 'failure event' is

$$
\begin{aligned}
F &= \{u < c\mathcal{L}(\theta)\} \\
&= \{c\mathcal{L}(\theta) - u > 0\} \\
&= \{Y > 0\}
\end{aligned}
$$

- Problem 1: It is not trivial to find an optimal c.
- Problem 2: c must be chosen from the beginning.

Modified BUS

- Instead of

$$F = \{u < c\mathcal{L}(\theta)\}$$

we could redefine

$$F = \left\{\frac{\mathcal{L}(\theta)}{u} > \frac{1}{c}\right\} = \left\{Y > \frac{1}{c}\right\}$$

- However, $Y = \frac{\mathcal{L}(\theta)}{u}$ is not a good choice since

$$\mathbb{E}[Y] = \mathbb{E}[\mathcal{L}(\theta)]\mathbb{E}[u^{-1}]$$

with $\mathbb{E}[u^{-1}] = \int_0^1 u^{-1} du = \ln u|_0^1$.

Modified BUS

- **Proposal** (DiazDelaO et al., 2017): Define

$$\begin{aligned} F &= \left\{ \ln\left[\frac{\mathcal{L}(\theta)}{u}\right] > -\ln c \right\} \\ &= \{Y > b\} \end{aligned}$$

Theorem

Let $\theta \in \mathbb{R}^n$ and $u \in \mathbb{R}$ be independent random variables such that $\theta \sim \pi(\theta)$ and $u \sim U[0,1]$. Let $\mathcal{L}(\theta)$ be a likelihood function and $\mathcal{D}$ a data set. Let $Y \equiv \ln\left[\frac{\mathcal{L}(\theta)}{u}\right]$. Thus, for all $b > b_{min} \in \mathbb{R}$:

1. $p(\theta|Y > b) = \pi(\theta|\mathcal{D})$
2. $\mathcal{P}(\mathcal{D}) \equiv \mathcal{P}_D = e^b \mathcal{P}(Y > b)$

Proof

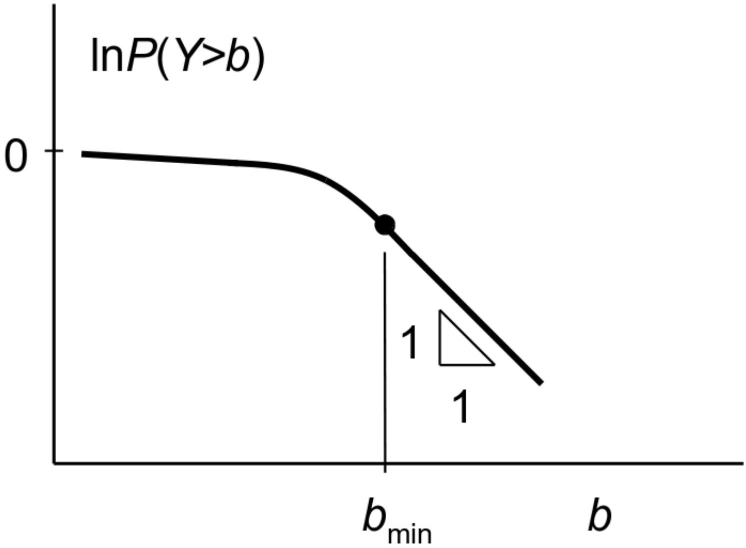
1. $p(\theta | Y > b) = \pi(\theta|\mathcal{D})$
   The proof is analogous to the previous rejection proof since

$$\mathcal{I}(Y > b) = \mathcal{I}(u < c\mathcal{L}(\theta))$$

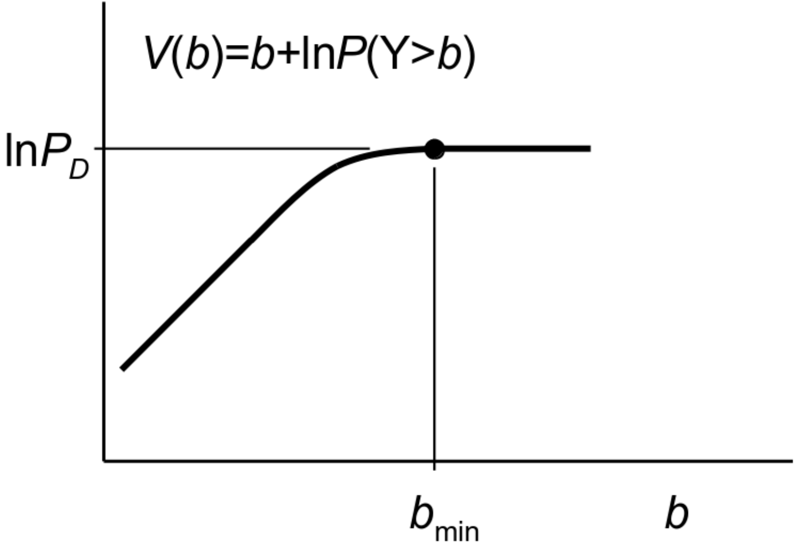
2. $\mathcal{P}(\mathcal{D}) \equiv \mathcal{P}_D = e^b \mathcal{P}(Y > b)$
   To prove this, note that

$$\begin{aligned} \mathcal{P}(Y > b) &= \int \int \pi(\theta)\mathcal{I}(0 \leq u \leq 1)\mathcal{I}(\ln \left[\frac{\mathcal{L}(\theta)}{u}\right] > b) du d\theta \\ &= \int \pi(\theta) \int_0^1 \mathcal{I}(u < e^{-b}\mathcal{L}(\theta)) du d\theta \\ &= e^{-b} \int \pi(\theta)\mathcal{L}(\theta) d\theta = e^{-b}\mathcal{P}_D \end{aligned}$$

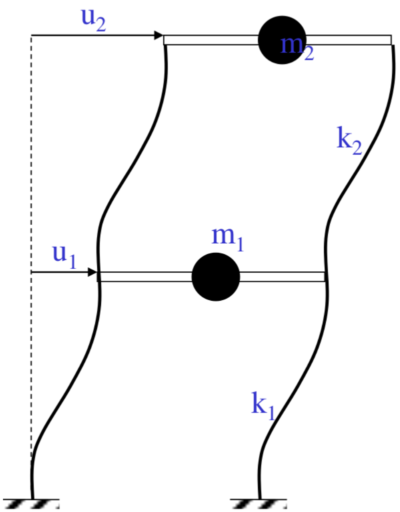
Characteristic Trends



Characteristic Trends



Example



- Consider the following two DOF shear building model.
- The stiffnesses are given by $\theta_1 \bar{k}_1$ and $\theta_2 \bar{k}_2$ for $\bar{k}_i = 29.7 \times 10^6$ N/m.
- The joint prior distribution for θ_1 y θ_2 is assumed to be the product of two lognormals with modes 1.3 and 0.8 and unit standard deviation.
- Our goal is to identify θ_1 and θ_2 given modal data.

Example

- The likelihood is given by

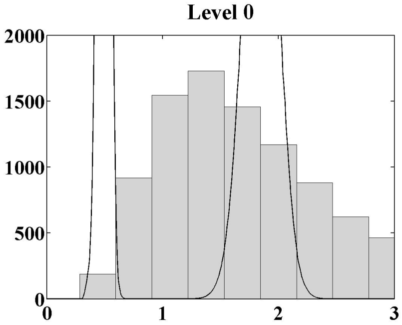
$$\mathcal{L}(\theta) \propto \exp\left\{-\frac{J(\theta)}{2\sigma_\epsilon^2}\right\}$$

  where

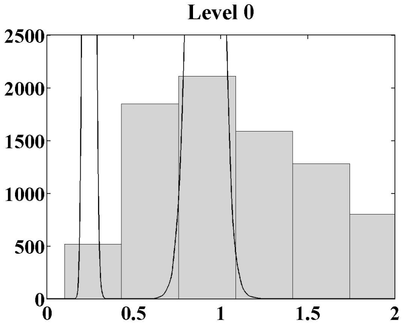
$$J(\theta) = \sum_{j=1}^{2} \lambda_j^2 \left[\frac{f_j^2}{\tilde{f}_j^2} - 1\right]^2$$

- $\mathcal{D} = \left\{\tilde{f}_1, \tilde{f}_2\right\} = \{3.13, 9.83\}\text{Hz}$
- f_1 y f_2 are natural frequencies obtained through finite element modelling.

Marginal Distributions

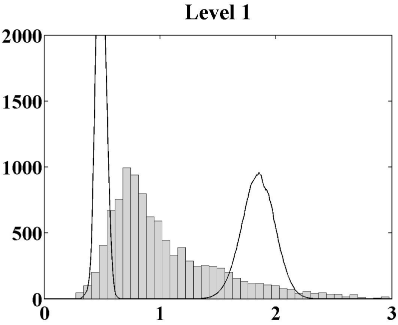


θ_1

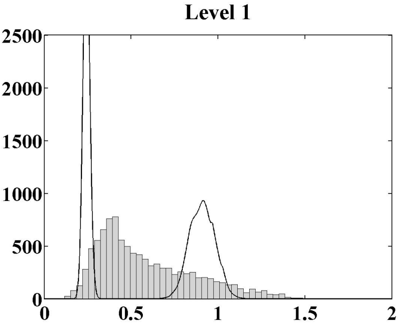


θ_2

Marginal Distributions

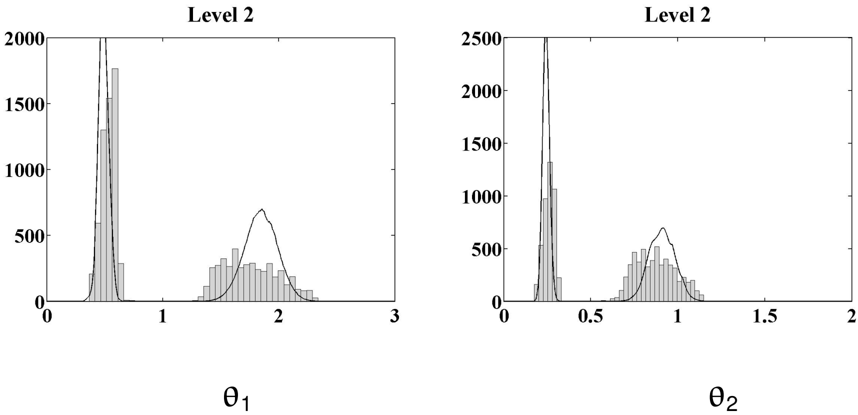


θ_1

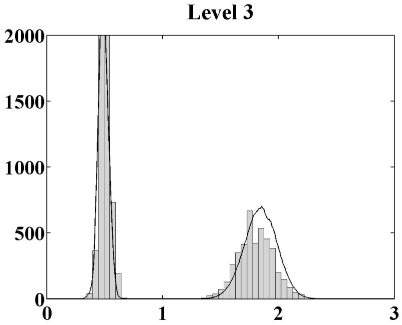


θ_2

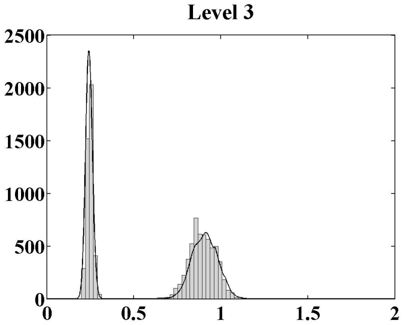
Marginal Distributions



Marginal Distributions

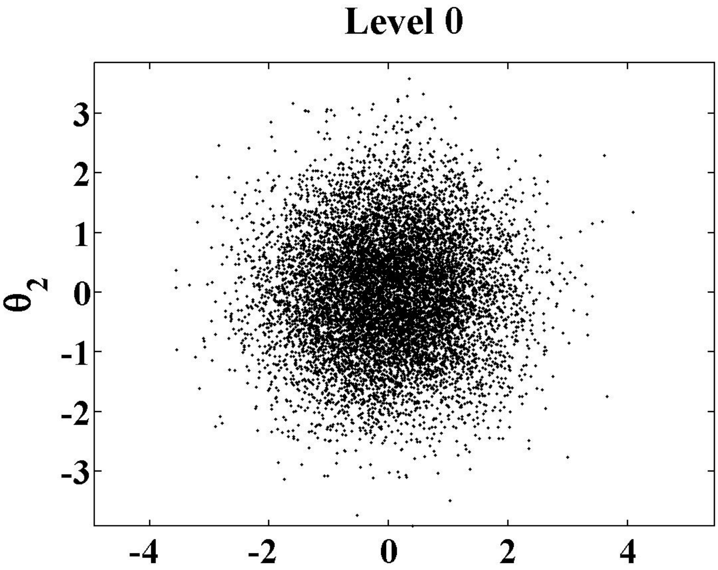


θ_1

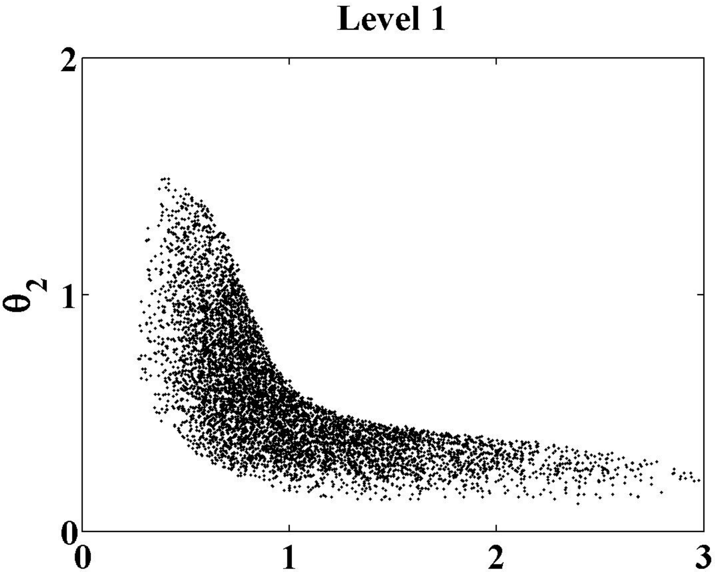


θ_2

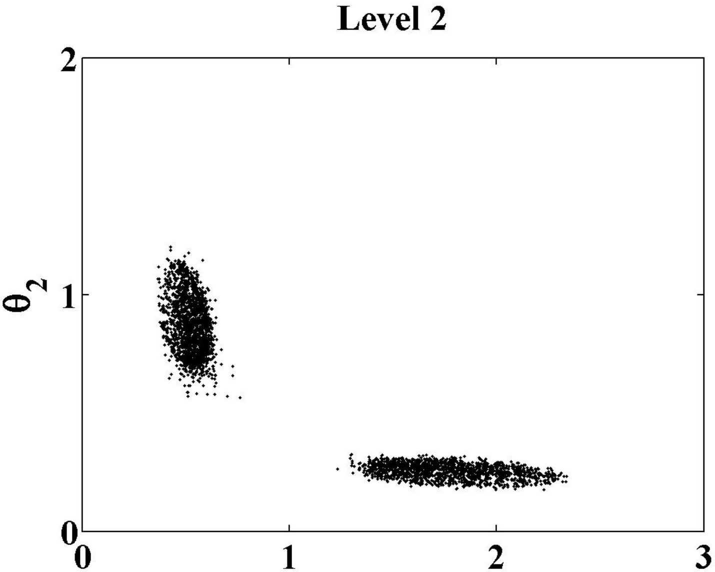
Parameter Space



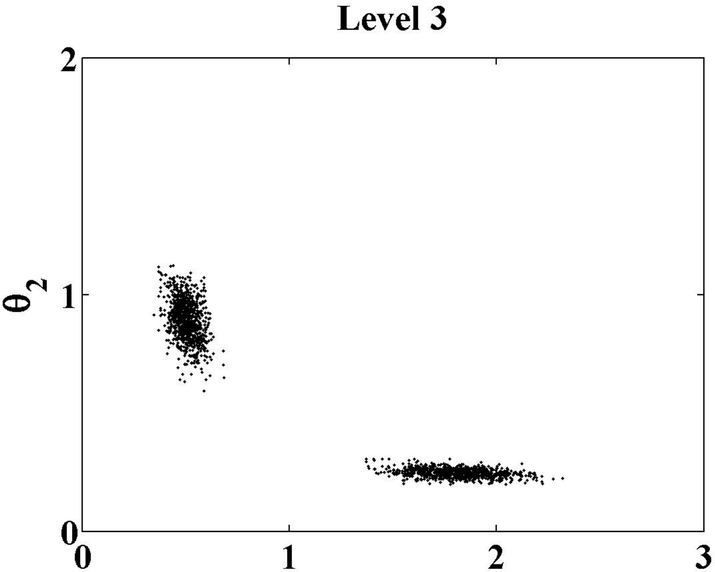
Parameter Space



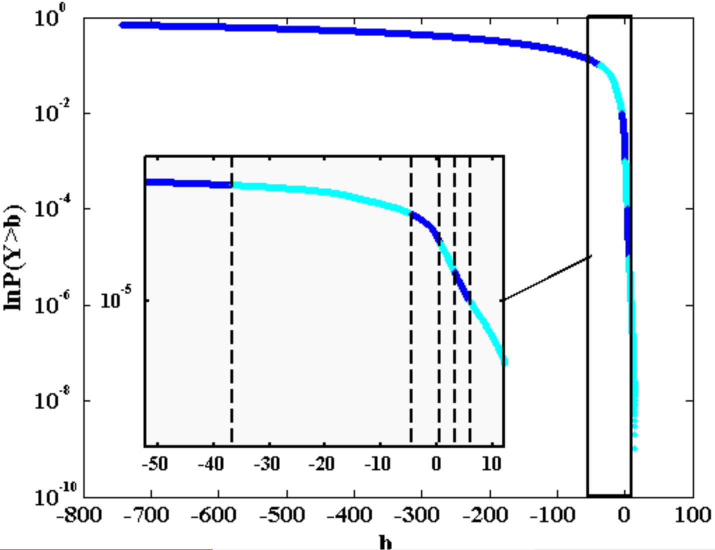
Parameter Space



Parameter Space

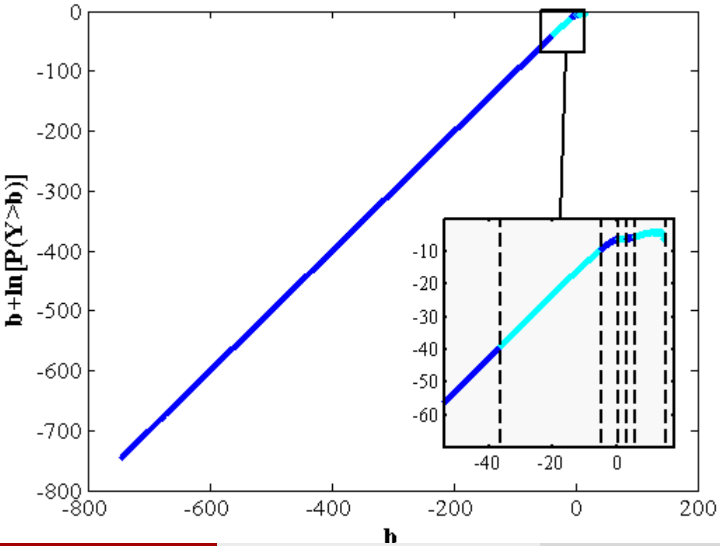


MBUS



lnP(Y>b)
b
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
-800
-700
-600
-500
-400
-300
-200
-100
0
100
10^{-5}
-50
-40
-30
-20
-10
0
10

MBUS



0
-100
-200
-300
-400
-500
-600
-700
-800
-800
-600
-400
-200
0
200
b+ln[P(Y>b)]
b
-10
-20
-30
-40
-50
-60
-40
-20
0

Automatic Stopping Criterion

- We know that
$$P_{F_k} = e^{-b_k} P_D$$
- Let $B_k = \{\theta : \mathcal{L}(\theta) > e^{b_k}\}$ be an inadmissible set. We can prove that
$$P_{F_k} = P_{\theta}\{B_k\} + e^{-b_k} P_D \, P_{\theta|D}\{B_k^c\}$$
- Moreover, we can prove that
$$\varnothing \subset \ldots \subset B_{k+1} \subset B_k,$$
- By defining $a_k = P_{\theta}\{B_k\}$ we have a monotone decreasing sequence of values such that $a_k \searrow 0$.
- Therefore, we can stop the algorithm for a small enough value of a_k which can be determined with an “outer” SuS run.

GPE Hyper-parameters

- Build the surrogate with data from training runs

$$\mathcal{D} = \{(y_1, \vec{x}_1), \ldots, (y_n, \vec{x}_n)\}.$$

- Assumed structure on the output

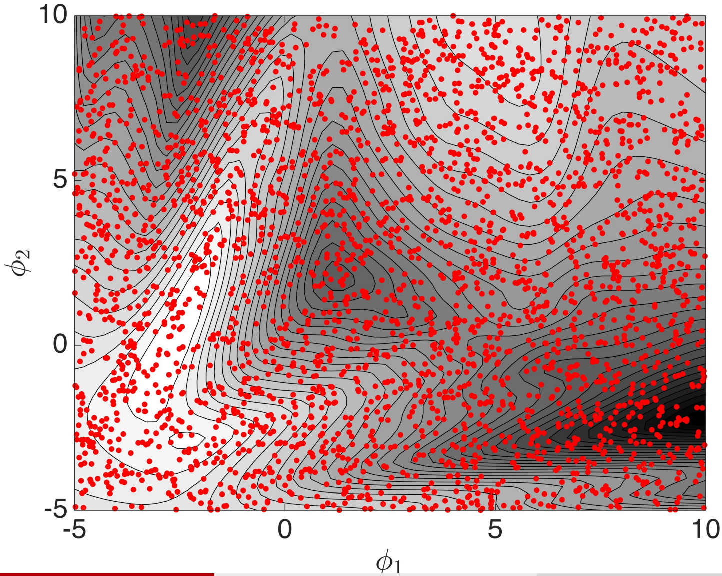
$$\eta(\vec{x}) = \underbrace{h(\vec{x})^T \beta}_{\text{Global trend}} + \underbrace{Z(\vec{x}|\sigma^2, \phi)}_{\text{Local variations}}$$

- Covariance function (kernel)

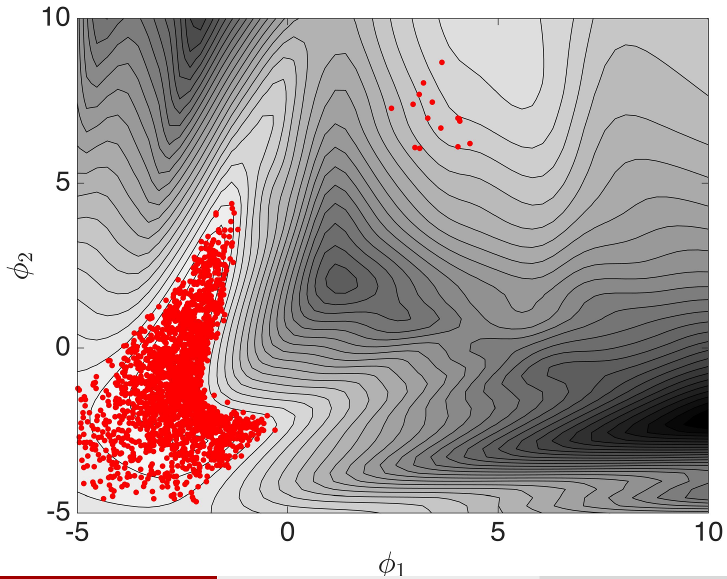
$$k(\vec{x}, \vec{x}'|\phi) = \sigma^2 \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \frac{(x_i - x_i')^2}{\phi_i} \right\}.$$

- Hyper-parameters: $\boldsymbol{\theta} = \{\beta, \sigma^2, \phi\}$

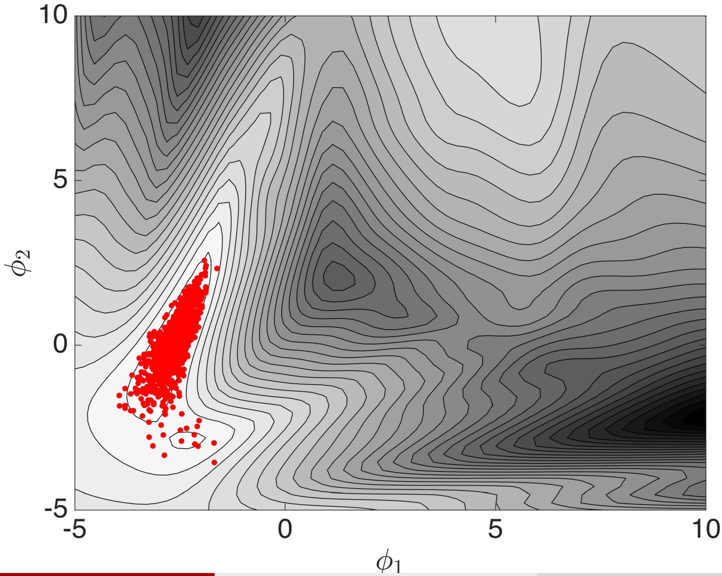
GPE Hyper-parameters



GPE Hyper-parameters



GPE Hyper-parameters



Summary

- BUS (Straub and Papaioannou, 2014) bridges the gap between the reliability problem and the Bayesian inference problem.
- As it is formulated, BUS requires the choice of a multiplier that, chosen incorrectly, the performance of the algorithm is affected.
- DiazDelaO et al. (2017) redefine the failure event, expressing the driving variable without the need of the multiplier.
- The implementation no longer requires a predetermined value of the multiplier, thus eliminating the need to rerun the algorithm in case an inadmissible or inefficient value is chosen.
- Different stopping criteria can be implemented.

References

1. Au, S.K. and Beck, J. (2001) Estimation of small failure probabilities in high dimensions by subset simulation, *Probabilistic Engineering Mechanics*, 16 (4), 263-277.

2. Straub, D. and Papaioannou, I. (2014) Bayesian updating with structural reliability methods, *Journal of Engineering Mechanics*.

3. DiazDelaO, F.A., Garbuno-Inigo, A., Au. S.K., Yoshida, I. (2017) Bayesian updating and model class selection with subset simulation, *CMAME*.
4. Zuev, K.M. (2015) Subset Simulation Method for Rare Event Estimation: An Introduction, *Encyclopedia of Earthquake Engineering*.

Thank you!