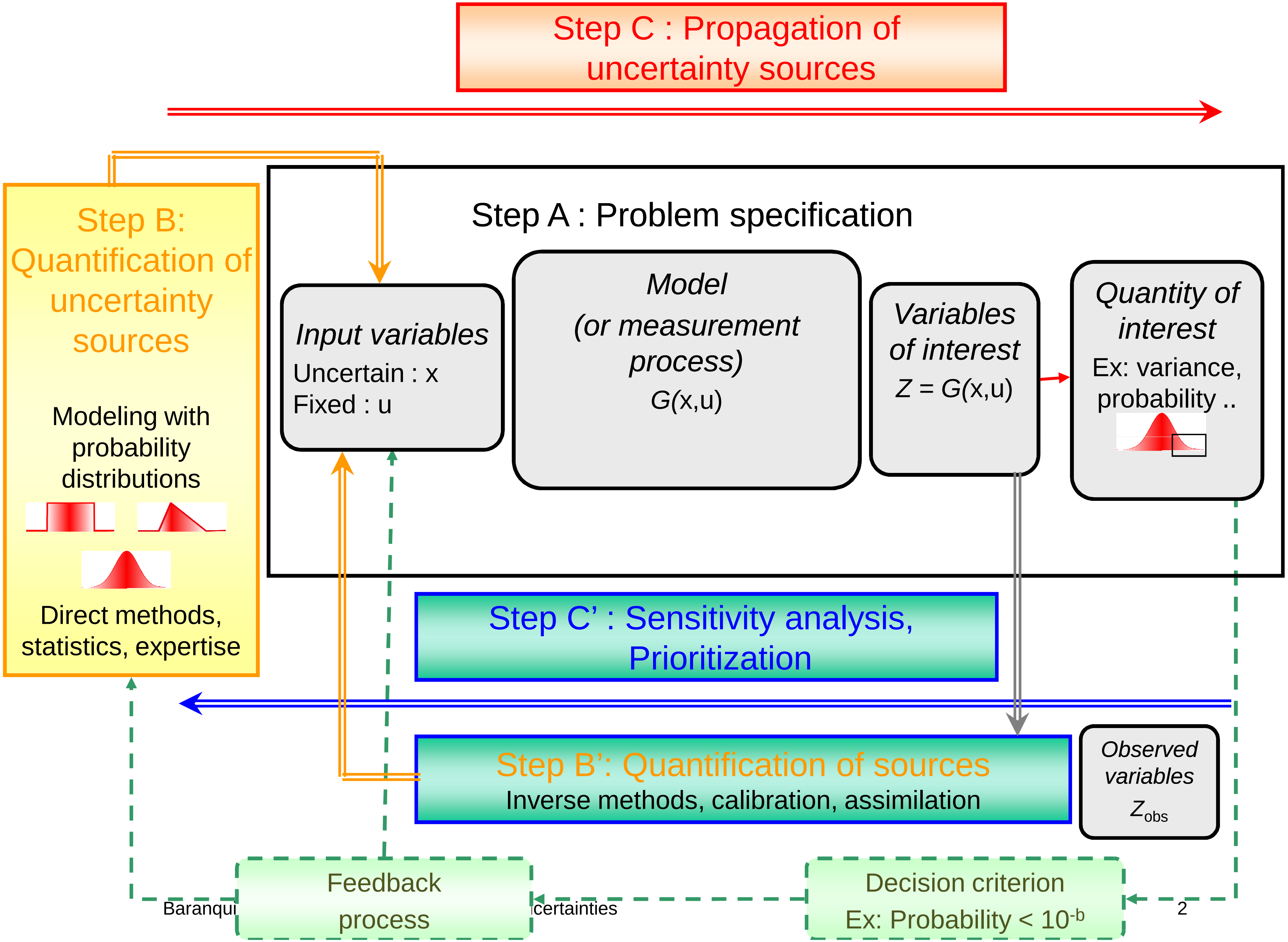
Propgation of uncertainties through numerical models

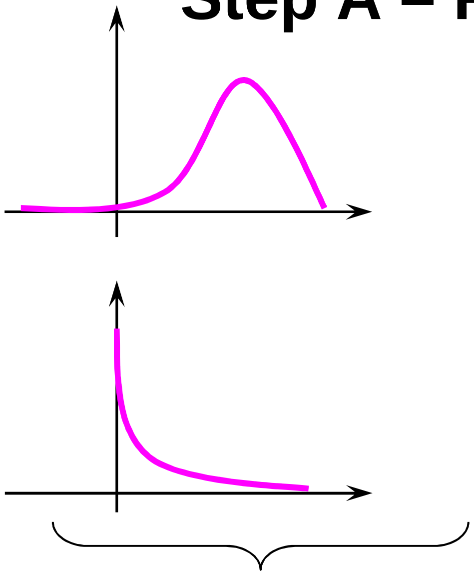
Fabrice Gamboa
Bertrand Iooss

28/02/2013

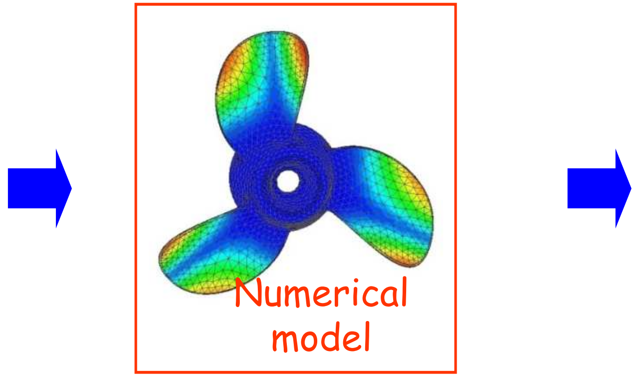
Uncertainty management - The generic methodology

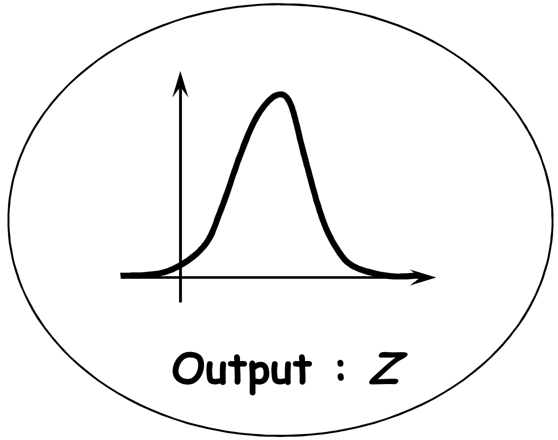


Step A – Focus on the quantity of interest

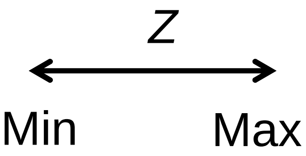


Inputs : X

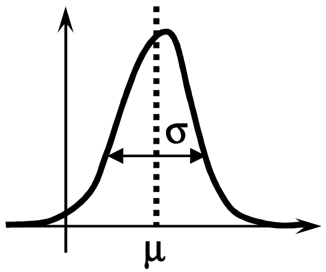




What is really interesting in our study?

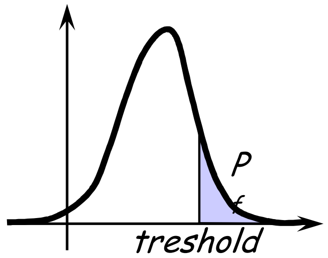


Ex: in the prior stage
of a new product design



Mean, median, variance,
(moments) of Z

Ex: in the design stage



Quantiles (extrems), probability
of treshold exceedence

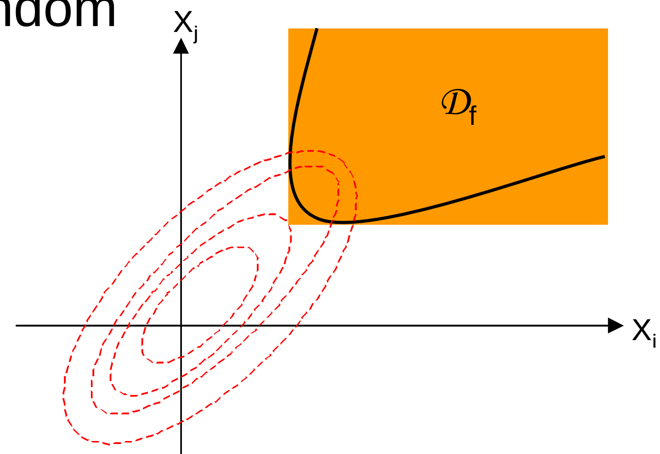
Ex: in the certification stage

Formally, the quantity of interest is a particular feature of the pdf of Z

A particular quantity of interest: the “probability of failure”

- G models a system (or a part of it) in operating conditions
  - Variable of interest $Z \rightarrow$ a given state variable of the system (e.g. a temperature, a deformation, a water level etc.)
- Following an « operator » point of view
  - The system is in safe operating condition if *Z* is above (or below) a given “safety” threshold
- System “failure” event: $Z \leq 0$
  - Classical formulation (no loss of generality) in which the threshold is 0 and the system fails when Z is negative
  - Structural Reliability Analysis (SRA) “vision”: Failure if *C-L* < 0 (Capacity – Load)
- Failure domain:
- Problem: estimating the mean of the random variable “

$$\mathcal{D}_f = \left\{ x \in \mathcal{X} : G(x) \leq 0 \right\}$$



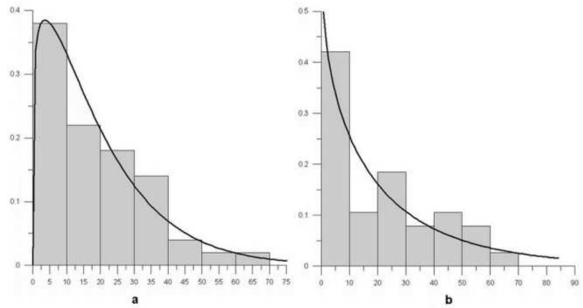
$$I_{\mathcal{D}_f}(x) = \mathbb{1}_{\{G(x) \leq 0\}}$$

$$p_f = \int_{\mathcal{D}_f} f(x)dx = \int_{\mathcal{X}} I_{\mathcal{D}_f}(x)\ f(x)dx = \mathbb{E}\left[I_{\mathcal{D}_f}(X)\right]$$

Step B - Quantification of uncertainty sources

Different cases with respect to available information

1. A lot of data
   - Fitting of probability distributions
   - Statistical hypothesis test (often parametric tests)



2. Few data (n < 10)
   - Hypothesis on parametric probability distribution
   - Non-parametric tests : less powerfull, wide bounds
   - Expert judgement, then Bayesian inference

3. No data
   - Expert judgment techniques
   - Maximum entropy principle

$$H(X) = -\int_{\mathcal{X}} f(x) \log\left(f(x)\right) dx$$

Measure of the "vagueness" of the information on X provided by f(x)

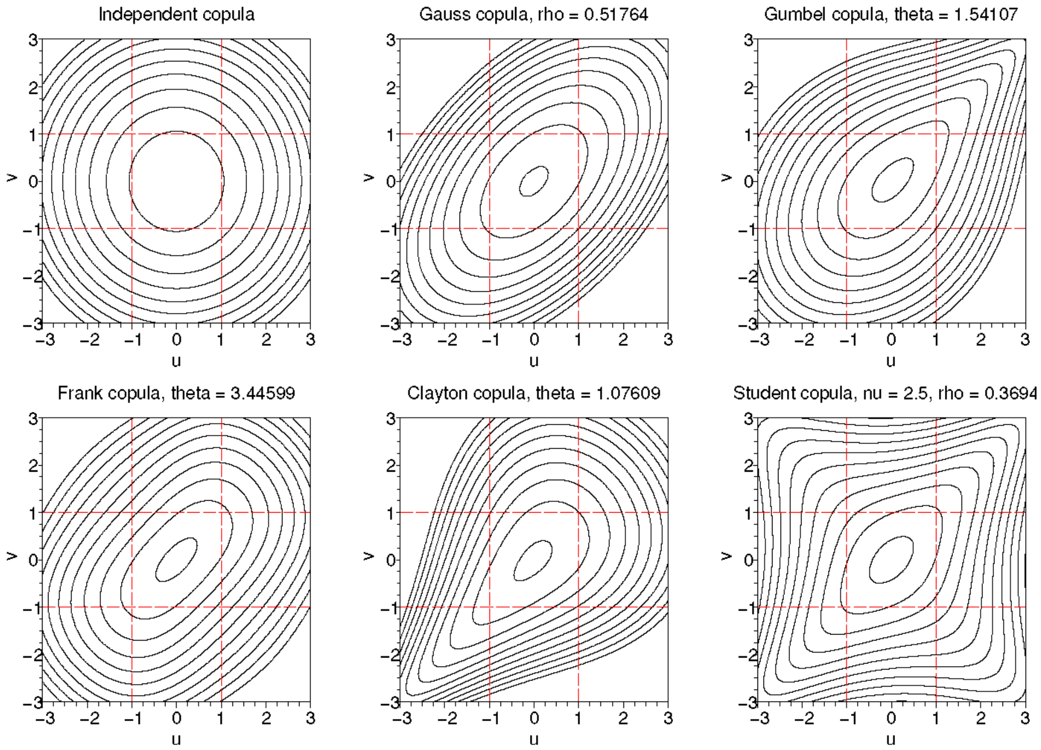
Information	Maximum Entropy pdf
$X \in [a,b]$	Uniform $X \sim \mathcal{U}(a,b)$
$\mathbb{E}(X) = \mu$ $X \in [0, \infty[$	Exponential $X \sim \mathcal{E}(1/\mu)$
$\mathbb{E}(X) = \mu$ $\mathbb{V}(X) = \sigma^2$	Normal $X \sim \mathcal{N}(\mu, \sigma)$

Some comments (Step B). Dependency

- Taking into account the dependency between inputs is a crucial issue in uncertainty analysis
  - Using copulas structure → CDF of the vector X as a function of the marginal CDF of $X_1 \dots X_n$:

$$F(x_1, x_2, \cdots, x_n) = C\left(F(x_1), F(x_2), \cdots, F(x_n)\right)$$

Example: All bivariate densities here have the same marginal pdf's (standard Normal) and the same Spearman rank coeff. (0.5)



Step C - Uncertainty propagation: main principles

Propagate uncertainties from X to *Z*, via the deterministic function *G* (•)

- Conceptually simple problem, but with sometimes a complex implementation
- Choice of method strongly depends on the quantity of interest

=> importance of step A

This quantity of interest is linked to decisional issues

<u>Two kinds of problems :</u>

- Central tendancy (ex. mean) or dispersion (variance)
  - Metrology

➡ Analytical methods sometimes applicable

- High quantile, « probability of failure »
  - → justification of a safety criterion

➡ Numerical methods (optimization, Monte Carlo sampling)

Step C' - Sensitivity analysis: main objectives

- **Reduction of the uncertainty of the model outputs by prioritization of the sources**

  - Variables to be fixed in order to obtain the **largest reduction** (or a fixed reduction) **of the output uncertainty**

    A purely mathematical variable ordering

  - Most influent variables in a given output domain

    - if reducibles, then R&D prioritization

    - else, modification of the system

The individual cost of the reduction may change the previous variable ordering

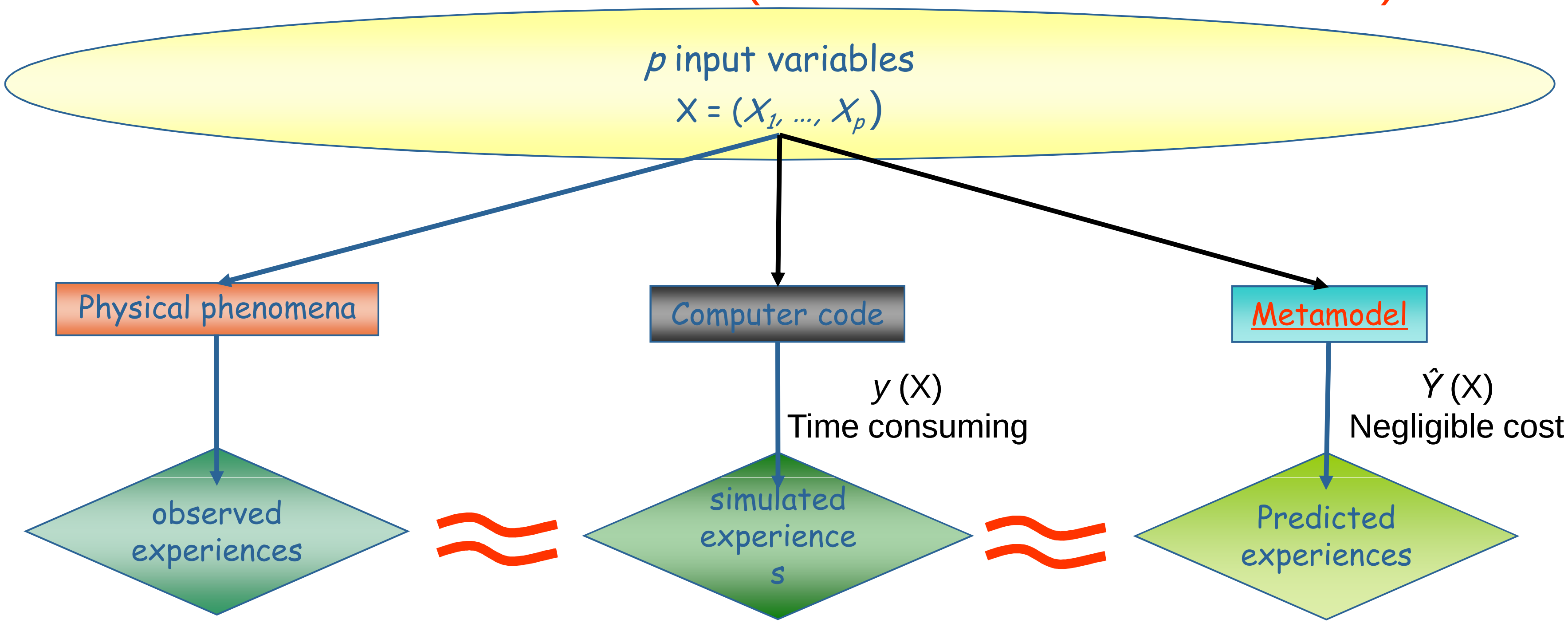
- **Simplification of a model**

  - **determination of the non-influent variables,** that can be fixed without consequences on the output uncertainty

  - building a simplified model, a metamodel

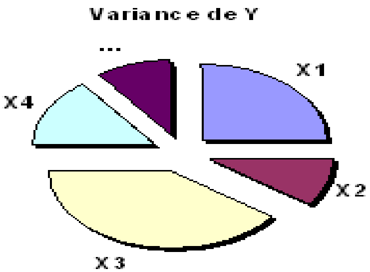
Uncertainties management for cpu time consuming models

A useful solution : the metamodel (model of the numerical model)

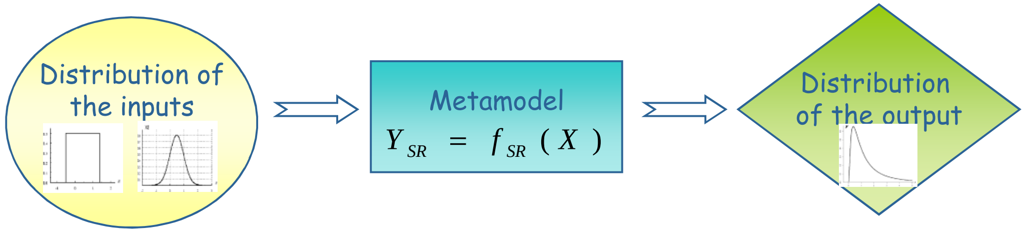


Use of the metamodel :

- C':Sensitivity analysis



- C:Uncertainty propagation (via Monte Carlo methods)



- B':Calibration

Identification of input parameters values

Adequation between observed and simulated experiences

V&V process: Verification and Validation

To sum up :

Verification : do I solve the equations right ?

Validation : do I solve the right equations ?
(at least for the intended application)

Two levels for Verification :

1. Code Verification : some kind of "*internal*" correctness of the code

   may be assessed by formal methods from Software Engineering

2. Calculation Verification : concerns the calculations themselves

   Convergence, grid adaptation, solution algorithms, …

   Is the solution closed to the exact one ?

We'll talk later on subtilties between Code and Model(s)

Some mathematical methods for uncertainty propagation

Quadratic combination method

Data : mean values of X_i : $\mu_i = \mathbb{E}[X_i]$

variance-covariance matrix of X_i :

$$Cov[X_i, X_j] = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$
$$\rho_{ij} = \mathbb{E}\left[\frac{X_i - \mu_i}{\sigma_i}\frac{X_i - \mu_i}{\sigma_j}\right]$$

Taylor expansion of G(•) around E(X) :

$$G(X) = G(\mu) + \sum_{i=1}^{N} \frac{\partial G}{\partial X_i}\Big|_{X=\mu}(X_i - \mu_i)$$
$$+ \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial^2 G}{\partial X_i \partial X_j}\Big|_{X=\mu}(X_i - \mu_i)(X_j - \mu_j) + o\left(\|X - \mu\|^2\right)$$

En général, dans les applications le développement est d'ordre 1

Quadratic combination method – First order

Mean of Z

$$\mathbb{E}\left[Z\right] = G(\mu)$$

Variance of Z

$$\mathbb{V}\left[Z\right] = \mathbb{E}\left[(Z - \mathbb{E}\left[Z\right])^2\right] = \mathbb{E}\left[\left(G(\mu) + \sum_{i=1}^{N} \frac{\partial G}{\partial X_i}\Big|_{X=\mu} (X_i - \mu_i) - G(\mu)\right)^2\right] =$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial G}{\partial X_i}\Big|_{X=\mu} \frac{\partial G}{\partial X_j}\Big|_{X=\mu} \mathbb{E}\left[(X_i - \mu_i)\,(X_j - \mu_j)\right]$$

$$\mathbb{V}\left[Z\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial G}{\partial X_i}\Big|_{X=\mu} \frac{\partial G}{\partial X_j}\Big|_{X=\mu} \rho_{ij}\sigma_i\sigma_j$$

Remarks :

++ *Needs only mean and covariance of X*

-- *Do not use if G(.) is strongly non linear*

-- *Provides only mean and variance of Z => no extrapolation fo the distrib. law of Z*

++ *if X is gaussian and G(.) is linear, then Z is gaussian*

Quadratic combination method – Independent case

If the X_i s are independent :

$$\mathbb{V}[Z] = \sum_{i=1}^{N} \left(\frac{\partial G}{\partial X_i}\Big|_{X=\mu} \right)^2 \sigma_i^2$$

Quadratic summation formula

Contribution of each input variable to the uncertainty of the output variable

$$\eta_i^2 = \frac{1}{\mathbb{V}[Z]} \left(\frac{\partial G}{\partial X_i}\Big|_{X=\mu} \right)^2 \sigma_i^2$$

Sensitivity indices (normed)

Methods of Monte Carlo simulation

– General mthods to evaluate a numerical quantity, using some random simulations

– In uncertainty propagation : use a random sample of G(X) to evaluate the quantity of interest

– We suppose that we know how to simulate an i.i.d (independent and identically distributed) sample of X_i following its probability distribution f_i

Monte Carlo in general (1/3)

- Computation of the integral :

$$I = \int_{\mathcal{X}} h(x) f(x) dx$$

h(•) : deterministic function
X : r.v. with densityf(x)

$$\int_{\mathcal{X}} h(x) f(x) dx = \mathbb{E}[h(X)]$$

$$x^{(1)}, x^{(2)}, ..., x^{(n)}$$

Random sample of X

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} h(x^{(i)}) \to \mathbb{E}[h(x)]$$

$$\hat{I} \to I$$

From the law of large numbers, the Monte Carlo estimator converges (a.s.) to the true quantity

Monte Carlo estimator

Monte Carlo (2/3)

- Variance of the Monte Carlo estimator

$$\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} h(X^{(i)})\right] = \frac{1}{n^2}\, n\mathbb{V}\left[h(X)\right] = \frac{1}{n}\mathbb{V}\left[h(X)\right]$$

Variance of the sum of n r.v. i.i.d.

– Variance of h(X) is given via its estimator :

$$\mathbb{V}\left[h(X)\right] \approx \frac{1}{n}\sum_{i=1}^{n}\left(h(x^{(i)}) - \hat{I}\right)^2$$

– General expression for the variance of the MC estimator

$$\mathbb{V}\left[\hat{I}\right] \approx \frac{1}{n^2}\sum_{i=1}^{n}\left(h(x^{(i)}) - \hat{I}\right)^2$$

– We note :

$$\sigma_{\hat{I}}^2 = \mathbb{V}\left[\hat{I}\right]$$

Monte Carlo (3/3)

- Asymptotic law of the estimator from Central Limit Theorem:

$$\frac{\sqrt{n}}{\sigma_{h(X)}}\left(\hat{I}-I\right) \sim \mathcal{N}(0,1) \quad \text{with} \quad \sigma_{h(X)} = \sqrt{\mathbb{V}[h(X)]}$$

Low convergence speed (in $1/\sqrt{n}$) but:

Independence with respect to the dimension of X and to the form of h(•)

Unbiased estimator

Precision only depends on n (then on the cpu time of h(.))

– Confidence intervals for the Monte Carlo error:

$$\epsilon_n = \hat{I} - I$$ ← Monte Carlo error

$$\epsilon_n \in \left[-q_{(1-\alpha/2)} \cdot \sigma_{\hat{I}},\ q_{(1-\alpha/2)} \cdot \sigma_{\hat{I}}\right]$$ ← Confidence interval with probability α

Quantiles of the standard normal law (↑ $q_{(1-\alpha/2)}$)

$$\sigma_{\hat{I}} = \frac{\sigma_{h(X)}}{\sqrt{n}}$$

Monte Carlo and uncertainty propagation

- Propagation of the uncertainties of X to Z=G(X)

$x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ ← n-sample of X

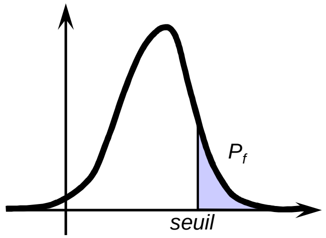
- Monte Carlo estimator of mean and variance of Z :

$$\mathbb{E}[G(X)] \approx \frac{1}{n} \sum_{i=1}^{n} G(x^{(i)})$$

$$\mathbb{V}[G(X)] \approx \frac{1}{n} \sum_{i=1}^{n} \left[G(x^{(i)}) - \frac{1}{n} \sum_{i=1}^{n} G(x^{(i)}) \right]^2$$

- Moments of Z are estimated by the empirical moments

Estimation of a probability of failure



- System failure : event $Z \leq 0$

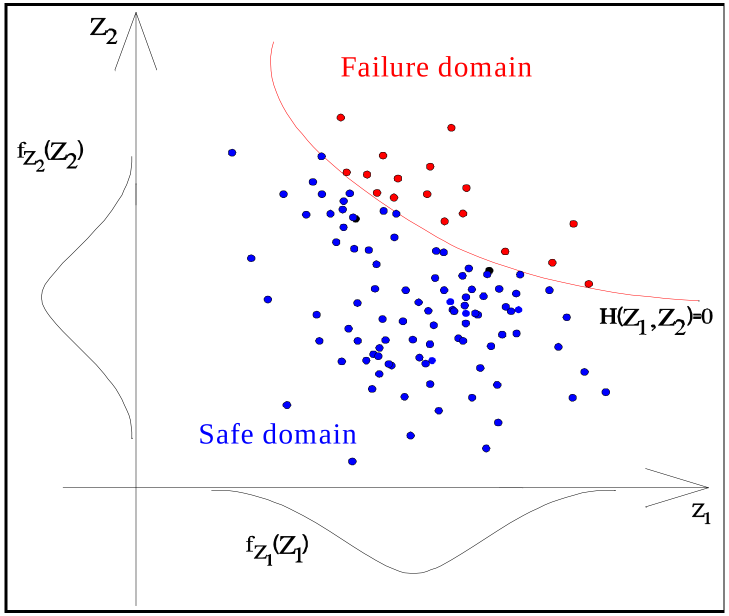
- Failure domain: $\mathcal{D}_f = \left\{ x \in \mathcal{X} : G(x) = z \leq 0 \right\}$

- Failure probability: $p_f = \int_{\mathcal{D}_f} f(x)dx = \int_{\mathcal{X}} I_{\mathcal{D}_f}(x)\ f(x)dx = \mathbb{E}\left[I_{\mathcal{D}_f}(X) \right]$

  – Problem : computation of the mean of the random variable $I_{\mathcal{D}_f}(x)$

- Failure indicator : $I_{\mathcal{D}_f}(x) = \mathbb{1}_{\{G(x) \leq 0\}}$

Monte Carlo estimation of p_f (1/3)



- Naive Monte Carlo estimator :

$$\hat{p}_f = \frac{1}{n} \sum_{i=1}^{n} I_{\mathcal{D}_f}\left(x^{(i)}\right)$$

- Variance of the estimator :

$$\mathbb{V}[\hat{p}_f] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} I_{\mathcal{D}_f}\left(x^{(i)}\right)\right] = \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} I_{\mathcal{D}_f}\left(x^{(i)}\right)\right]$$

- As $I_{\mathcal{D}_f}\left(X^{(1)}\right),\ I_{\mathcal{D}_f}\left(X^{(2)}\right),\ldots I_{\mathcal{D}_f}\left(X^{(n)}\right) \sim \mathcal{B}(p_f)$ Bernouilli $i.i.d.$

- We have: $\mathbb{V}[\hat{p}_f] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left[I_{\mathcal{D}_f}(x)\right] = \frac{1}{n^2}\ n\ p_f(1-p_f)$

$$\mathbb{V}[\hat{p}_f] = \frac{1}{n} p_f(1-p_f)$$

Estimated by:

$$\mathbb{V}[\hat{p}_f] \approx \frac{1}{n}\hat{p}_f(1-\hat{p}_f)$$

- Asymptotical convergence to a normal law and other MC estimator properties

Monte Carlo estimation of p_f (2/3)

- Decrease in square root of n :

$$\sigma_{\hat{p}_f} = \frac{1}{\sqrt{n}}\sqrt{p_f(1-p_f)}$$

- Variation coefficient :

$$cv = \frac{\sigma_{\hat{p}_f}}{\mathbb{E}[\hat{p}_f]} = \sqrt{\frac{p_f(1-p_f)}{n}\frac{1}{p_f^2}} = \sqrt{\frac{1-p_f}{n\ p_f}}$$

- For small values of p_f :

$$p_f \to 0 \implies \frac{1-p_f}{p_f} \to \frac{1}{p_f}$$

$$cv \approx \sqrt{\frac{1}{n\ p_f}}$$

« Relative error », precision of the estimate

- For example, if we estimate a proba $p_f = 10^{-r}$ with cv = 10%,

$$\sqrt{\frac{1}{n\ 10^{-r}}} = 10^{-1} \implies n = 10^{r+2}$$

10^{r+2} valuess of G(X), then 10^{r+2} calls to the code G !

⇒prohibitive required cpu times

⇒Use of improved methods: approximate methodes (FORM/SORM), accelerated Monte Carlo methods, metamodel-based methods, …

Importance sampling (1/3)

- Idée : modifying the sampling prob. distribution of X in order to concentrate the samples in most interesting regions (in terms of contribution to computation of expectation of h(X))

$$I = \int_{\mathcal{X}} h(x) f(x) dx = \int_{\mathcal{X}} h(x) \frac{f(x)}{\varphi(x)} \varphi(x) dx = \int_{\mathcal{X}} h(x) w(x) \varphi(x) dx$$

- It's the expectation of the function $h(x)w(x), \quad X \sim \varphi(x)dx$

1) Produce a sample ($x^{(i)}$) from density $\varphi(x)dx$

2) Then, compute : $\hat{I}^{is} = \frac{1}{n} \sum_{i=1}^{n} h(x^{(i)}) w(x^{(i)})$

$$\mathbb{V}\left[\hat{I}^{is}\right] = \frac{1}{n} \mathbb{V}\left[h(X) \frac{f(x)}{\varphi(x)}\right]$$

- Unbiased estimator of I, in condition that the support of φ(x) contains the one of f(x)

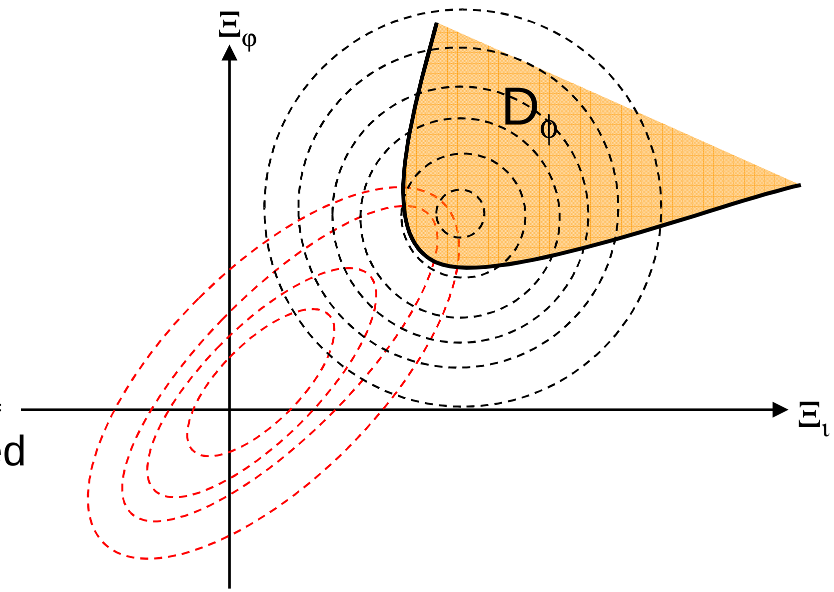
Importance sampling (2/3)

- This method does not guarantee any variance reduction ∀ φ(x)
- The choice of the « instrumental law » φ(x) is crucial
  - theoretically: optimal density : $$\varphi^*(x) = \frac{|\ h(x)|\ f(x)}{\int_{\mathcal{X}} |\ h(x)|\ f(x) dx}$$
  - The normalization constant is as difficult to evaluate as evaluating I !
  - However, practical result …
- Estimation of a failure probability p_f by importance sampling
  - Here : $h(x) = I_{\mathcal{D}_f}(x) = \mathbb{1}_{\{G(x) \leq 0\}}$
  - Optimal density : $$\varphi^*(x) = \frac{I_{\mathcal{D}_f}(x) f(x)}{\int_{\mathcal{X}} I_{\mathcal{D}_f}(x) f(x) dx} = \frac{I_{\mathcal{D}_f}(x) f(x)}{p_f}$$

Importance sampling (3/3)

- The optimal density id the conditional law of de X knowing that $X \in D_f$
- Intuitive result→ the method is mostly efficient if it produces samples in the failure domain

- Some practical algorithms:
  - Obtain a first idea of the configuration of D_f (first Monte Carlo runs)
  - Center the instrumental law on a point of D_f (for example on the design point P* obtained with the FORM method)



FORM method (1/5)

- FORM: First Order Reliability Method

- From structural safety domain

- 3 steps :

1. Transformation of inputs X_i to other inputs whose probability distributions have « good properties »

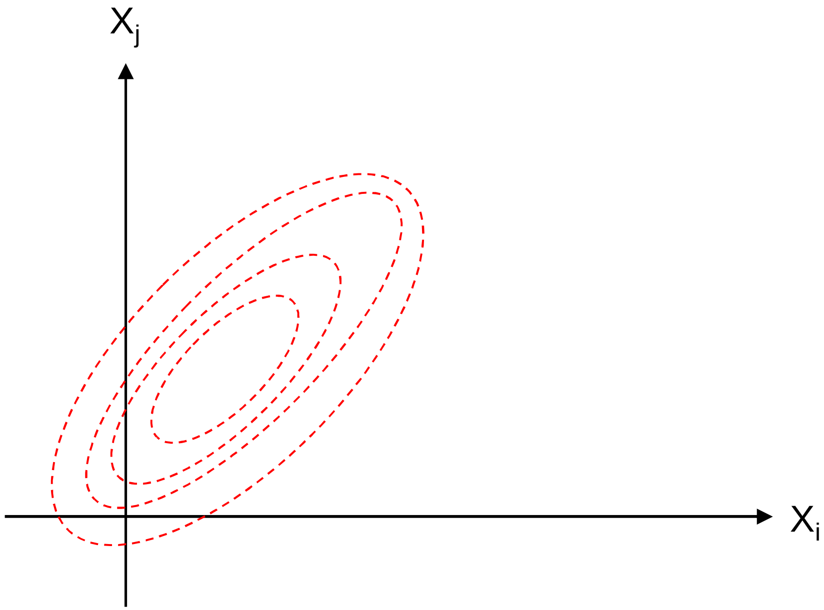
   Isoprobabilistic transformation → standard Gaussian space

2. Search of the most probable failure conditions

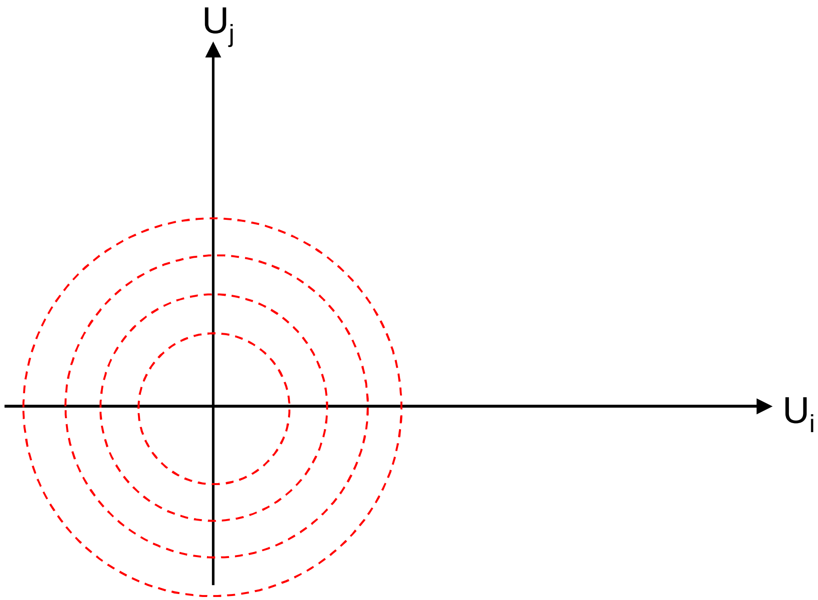
3. Estimation of the failure probability

FORM (2/5)
Isoprobabilistic transformation

$$U = \mathcal{T}(X)$$



Physical space



« Standard » space

- Each component of <u>U</u> follows a N(0,1)
- Component of <u>U</u> are independent
- The iso-prob. surfaces are spheres

$$\phi_d(\mathbf{u}) = \frac{1}{\left(\sqrt{2\pi}\right)^d} \exp\left(-\frac{1}{2} \sum_{i=1}^{d} u_i^2 \right)$$

Points which mostly contribute to p_f are the nearest to the origin in the standard space

FORM : isoprobabilistic transformation

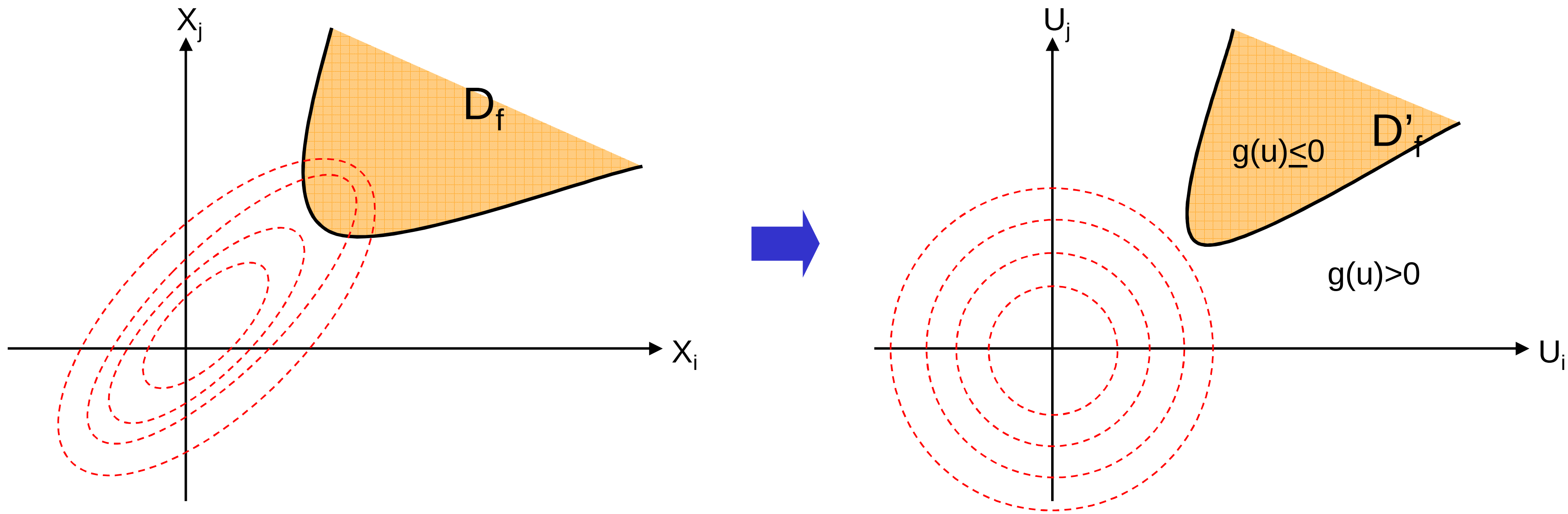
- Rosenblatt transformation

$$
T \quad : \quad
\begin{aligned}
u_1 &= \Phi^{-1}\left(F_1(z_1)\right) \\
u_2 &= \Phi^{-1}\left(F_2\left(z_2 \middle| z_1\right)\right) \\
&\vdots \\
u_N &= \Phi^{-1}\left(F_N\left(z_N \middle| z_{N-1}, \cdots, z_2, z_1\right)\right)
\end{aligned}
$$

FORM (3/5) – Isoprobabilistic transformation

- New expression of the failure probability

$$\mathbb{P}\left(G(X) \leq 0\right) = \mathbb{P}\left(G(\mathcal{T}^{-1}(U) \leq 0\right) = \mathbb{P}\left(g(U) \leq 0\right)$$



Physical space

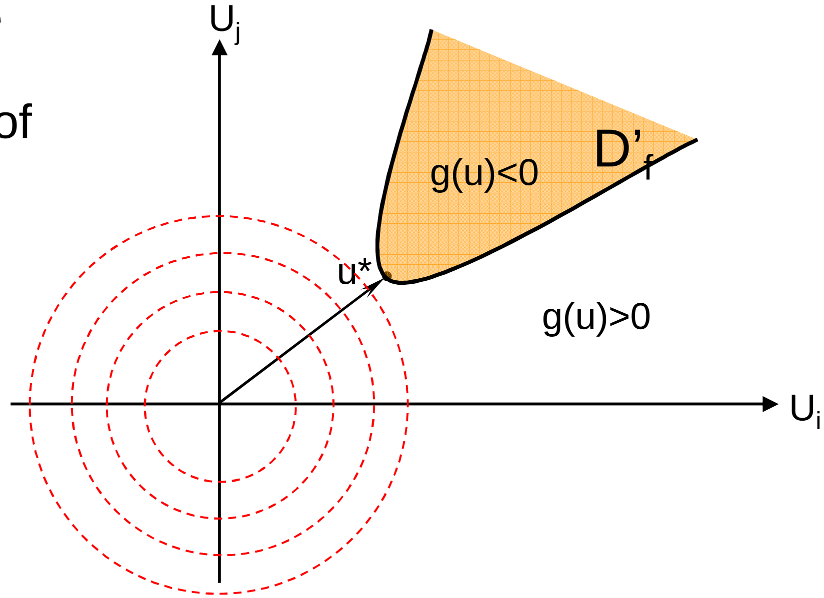
« Standard » space

– Expression of the failure probability :

$$p_f = \int_{\mathcal{T}(\mathcal{X})} \mathbb{1}_{g(u) \leq 0} f_u(u) du$$

FORM (4/5) – Search of the most probable failure conditions

- To each point of the standard space, some operating or failure conditions are associated
  - The most probable failure point is the nearest from the origin (where the desnity is max because the mean of of U is the null vector)
  - Reminder : the value of density $f_U(u)$ depends only on ||u|| (distance of u from origin)
  - Call it P* : designpoint
  - Call u* the vector OP*
  - The search of u*, (with an uniqueness hyp.), is an optimization problem under constraints



$$u^* = \min_{g(u)=0} \beta(u) = \min_{g(u)=0} \sqrt{u^t u}$$

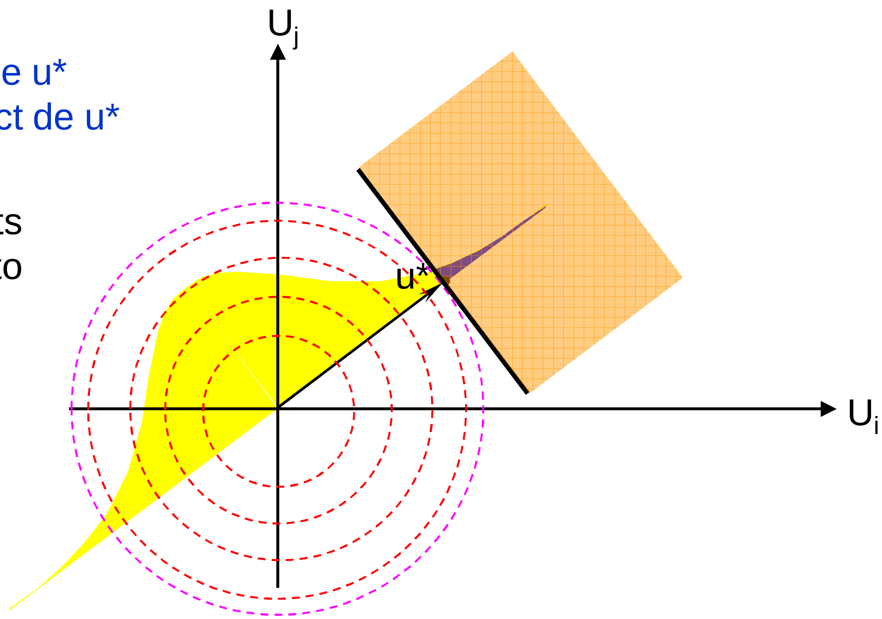
FORM (5/5) – Evaluation of p_f

- Hypothesis :
  - Replacement of the limit surface g(u)=0 by the hyperplan intersecting P* and orthogonal to u*, with equation :

$$\sum_{i=1}^{N} \alpha_i u_i + \beta = 0$$

β : norme de u*
α_i cos. direct de u*

  - Approximation based on the hyp. that points far away from P* have small contributions to p_f → their proba is very small



$$p_f \approx \mathbb{P}\left(\sum_{i=1}^{N} \alpha_i u_i + \beta \leq 0\right) = \mathbb{P}\left(\sum_{i=1}^{N} \alpha_i u_i \leq -\beta\right) = \Phi(-\beta)$$

Linear combination of r.v. N(0,1) with normed coeff. αi => N(0,1)

Distribution function of N(0,1)

β : « Reliability index »
α_i : « Importance factors» FORM → sensitivity indices of variables Ui to pf

FORM/SORM : Pros and cons

- Pros:
  - **reduced computing times** with respect to other methods
  - No dependancy between computing times and value of pf
  - Getting the importance factors and a design point

- Cons:
  - Approximation not always valid
  - No measure of the error which is made:
  - G has to be *differentiable*
  - Hypothesis of a unique design point

Quantile estimation (1/2)

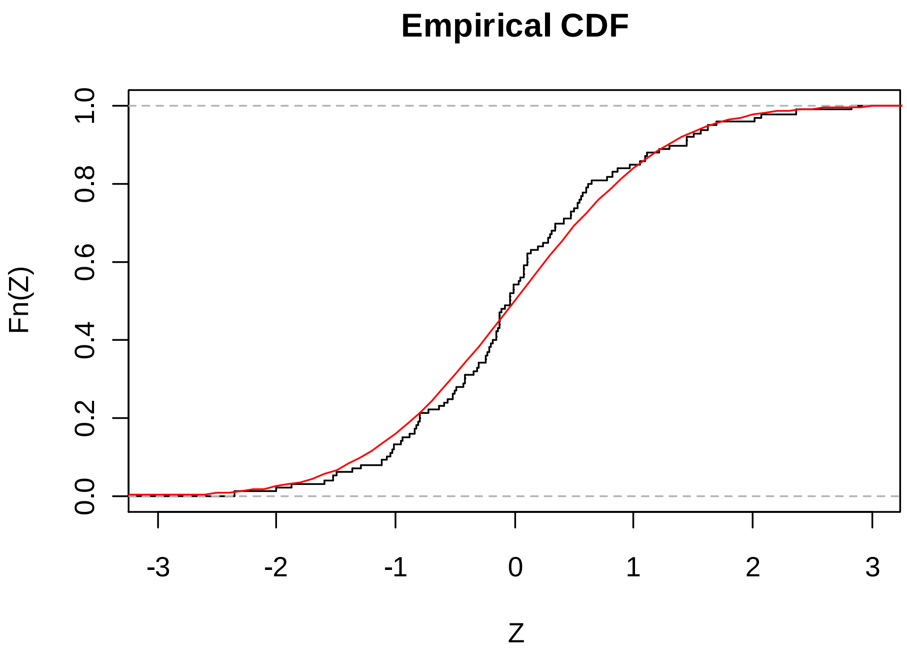
- Probability distribution function and quantile estimator

$$\hat{F}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{G(x^{(i)})\leq z\}}$$

← Definition

$$\hat{F}_n(z) \to F(z)$$

← Glivenko – Cantelli theorem: convergence to F(z)



Monte Carlo estimator of a p-quantile : empirical quantile

$$\hat{z}_p = \inf\left(z : \hat{F}_n(z) \geq p\right)$$

Quantile estimation (2/2)

- In practice:
  - Build an <u>ordered</u> sample from $G\left(x^{(1)}\right), G\left(x^{(2)}\right) ..., G\left(x^{(n)}\right)$
  - Call it: $z^{(1)}, z^{(2)}, ..., z^{(n)}$ $\quad z^{(1)} \leq z^{(2)} \leq ... \leq z^{(n)}$
  - $\hat{z}_p = z^{(\lceil np \rceil)}$
    - For example, if n=100 and p=0.95, then we have to take 96th value in the ordered sample
    - Of course, we need $\dfrac{1}{N} < p < 1 - \dfrac{1}{N}$

- Asymptotic law of the estimator:

$$\frac{\sqrt{n}}{\tau}\left(\hat{z}_p - z_p\right) \sim \mathcal{N}(0,1) \qquad \tau^2 = \frac{p(1-p)}{\left(f\left(z_p\right)\right)^2}$$

Quantile estimation – Wilks formula

- We can show that:

$$\mathbb{P}\left(z^{(np+r)} > z_p\right) = \sum_{j=n(1-p)-r+1}^{n} \mathbb{P}\left(j \text{ parmi les } z^{(i)} \text{ sont } > z_p\right) = 1 - C_p(n,r)$$

$$C_p(n,r) = \sum_{j=0}^{n(1-p)-r} \binom{n}{j} (1-p)^j \, p^{n-j}$$

  – Then, if r is the smallest integer such that $C_p(n,r) \leq 1-\beta \implies 1 - C_p(n,r) \geq \beta$

  – then, $\mathbb{P}\left(z^{(np+r)} > z_p\right) \geq \beta$

- We obtain the Wilks method
  – Conservative estimator of quantiles :
    - With a fixed n, find β (the confidence level of the quantile)
    - With a fixed β fixé, find n (required number of code runs)

Sampling via Wilks formula

Comments:

- Robust method
- No hypothesis on the distribution function

Constraint :

Can only be applied to pure random sample (i.i.d.)

Example with Wilks at first order and unilateral quantile

Z_{max} is the maximal value of the N-sample (i.i.d) of Z

$$P\left[P\left(Z \leq Z_{\max}\right) \geq \alpha\right] \geq \beta, \qquad N \text{ solution of } 1-\alpha^N \geq \beta$$

α	0.50	0.90	0.90	0.95
β	0.95	0.90	0.95	0.90
n	5	22	29	45

Application example

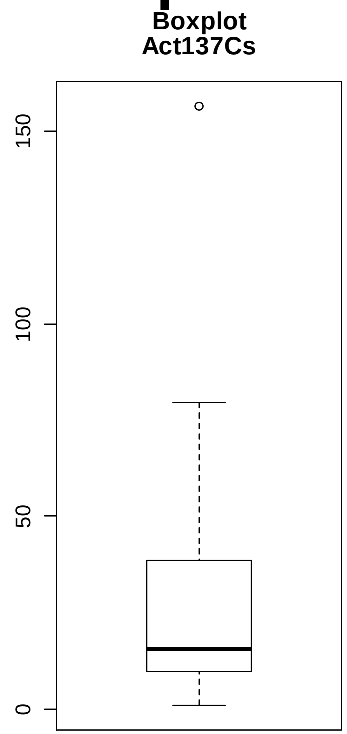
Example: pollution measures on a few points of a room wall

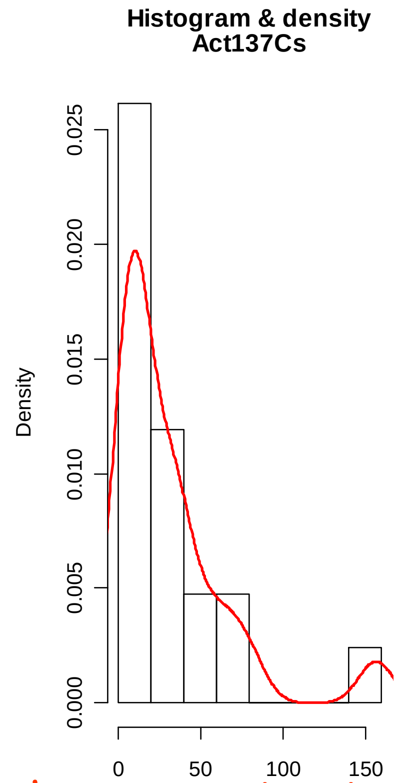
21 activity measures in ^{137}Cs (Bq/cm^2)

- Mean = 31.45
- Median = 15.4

- Standard-deviation = 36.11
- Min = 0.83 – Max = 156.67

- Skewness = 2.02
- Kurtosis = 4.19





Safety issue: guarantee (with a certain confidence level) that the contamination does not exceed a treshold over all the room wall

Examples : prediction of the amount of different category of wastes (proportion of activities < 50 Bq / cm^2 , > 100 Bq / cm^2 , …)

→ *waste quantities in different types of storage (deep geologic, subsurface, no storage)*

→ *different costs*

Useful probabilistic tools: universal inequalities

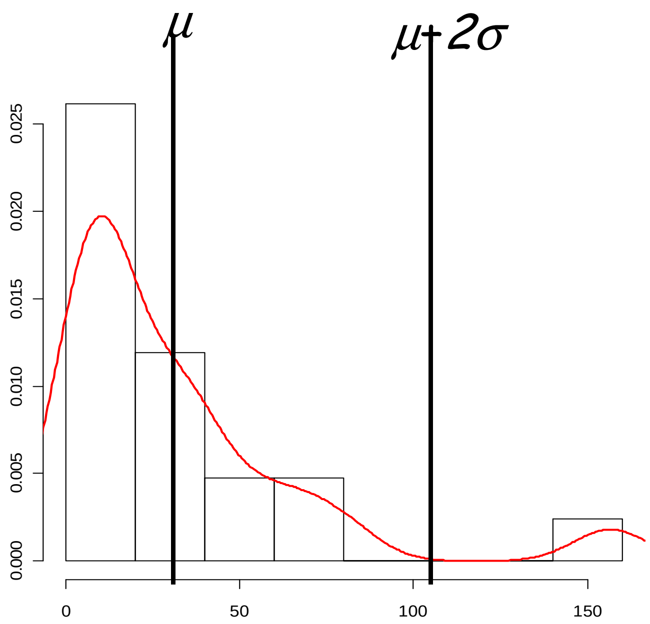
For a random variable X with mean μ and variance σ^2, we can use for $X > \mu$:

• the Bienaymé-Tchebytcheff inequality: $P(X \leq \mu + k\sigma) > \frac{k^2}{1+k^2}$

More than 72% of the surface < 100 Bq/cm²

Pessimistic bound

μ and σ^2 are replaced by their empirical estimates



• the Guttman inequality:

$$P(X \leq \mu + k\sigma) > \frac{q^2}{1+q^2} \text{ with } q^2 = \frac{(k^2-1)^2}{\gamma_2 - 1}$$

More than 82% of the surface < 100 Bq/cm²

Needs the knowledge of the kurtosis

• the Meidell inequality (*unimodality hypothesis*):

$$P(X \leq \mu + k\sigma) > \frac{(3k/2)^2}{1+(3k/2)^2}$$

More than 89% of the surface < 100 Bq/cm²

All these tools give unsafe estimates

Using the Wilks formula

For an i.i.d. sample $\{X_1,\dots,X_n\}$ of a random variable X, if n is solution of $1-\alpha^n \geq \beta$ and $X_{max} = \max\{X_1,\dots,X_n\}$ we have $$P\left[P\left(X \leq X_{\max} \middle| (X_1,\dots,X_n)\right) \geq \alpha\right] \geq \beta$$

It gives:

1. the minimal sample size n for α and β
2. for a given sample, the α-quantile value, with a β confidence degree

α	0.50	0.90	0.95	0.95
β	0.95	0.95	0.90	0.95
n	5	29	45	59

No hypothesis on distribution function and no needs of parameter estimates

More general result linking n and order r (rank in the ordered sample $\{X_{(1)},\dots,X_{(n)}\}$)

Application (measures in ^{137}Cs) :

- Wilks (n=21, r=2, β=0.9) -> more than 83% of the surface < 80 Bq/cm²
(with a 90% degree of confidence)
- Meidell (unimod., σ estimate) -> more than 80% of the surface < 80 Bq/cm²

Conclusions on step C (uncertainty propagation)

- **Challenge:** balance between precision of the estimate and cpu time cost

- Use **Monte Carlo** if possible: independent of input dimension, unbiased estimation, gives a confidence interval

  BUT : needs large number of model runs to obtain convergence

- If this cost is unreachable, alternative methods exist:
  - Accelerated Monte Carlo method (importance sampling, etc.)
  - Méthodes quasi-Monte Carlo (cf. cours 2) - But: curse of dimensionality
  - Approximate methids :
    - Quadratique summation - But: linear hypothesis
    - FORM/SORM : fast estimation of p_f . Can be used to initialize another method (importance sampling)

  - Using a surrogate model of the computer code (metamodel)

Bibliography

– David, H. A. (1981). *Order Statistics. Wiley, New York.*

– De Rocquigny, Devictor & Tarantola (eds), Uncertainty in industrial practice, Wiley, 2008

– M. Lemaire, A. Chateauneuf, and J. Mitteau, Structural reliability, Wiley Online Library, 2009

– Robert & Casella, Monte Carlo Statistical Methods, 2nd Edition, Springer, 2004

– Rubinstein, Simulation and the Monte Carlo method, Wiley, 1981

– B. Sudret, *Uncertainty propagation and sensitivity analysis in mechanical models*, HDR de l'Université Clermont II, 2008

– Guide to the expression of Uncertainty in Measurements (GUM), ISO publication