

IRSN

INSTITUT
DE RADIOPROTECTION
ET DE SÛRETÉ NUCLÉAIRE

Méthodes statistiques pour la prise en compte des incertitudes dans les modèles complexes

J. Baccou

Ecole d'été du GdR ModMat, 20-24/07/15

Plan du cours

1) Introduction

2) Analyse d'incertitude

3) Analyse de sensibilité

1) Introduction

Incertitude (de mesure) (GUM, JCGM 100:2008(F))

paramètre, associé au résultat d'un mesurage, qui caractérise la dispersion des valeurs qui pourraient raisonnablement être attribuées au mesurande

La prise en compte des incertitudes est transverse à tous les domaines de l'industrie pour de multiples raisons:

Raisons économiques :

Utilisation de modèles et de données réalistes permet de réduire des marges prises à cause d'hypothèses pénalisantes

Raisons de sûreté :

En France, l'article 4 de la loi sur la prévention des risques technologiques du 30 juillet 2003: « Le demandeur d'une installation nouvelle ou de la modification d'une installation existante doit fournir une étude de dangers ...

Cette étude donne lieu à une analyse de risque qui prend en compte la probabilité d'occurrence, suivant une méthodologie qu'elle explicite. Elle définit et justifie les mesures propres à réduire la probabilité et les effets de ces accidents».

Outil de communication :

Permet de rendre visible l'état des connaissances et donc d'expliquer les règles de sûreté prises.

...

- Codes de calculs simulant des phénomènes complexes
 - Approximation de la réalité physique
 - Données du code connues avec une certaine précision ou mal connues

 **La prédiction des codes n'est pas une valeur exacte**

- Pour gérer les incertitudes:
 - Codes conservatifs: « pire des cas »
 - Codes Best-Estimate: tenir compte des incertitudes

Formalisation de la démarche « Best-Estimate Code Plus Uncertainty »

Ex: CSAU (1989)

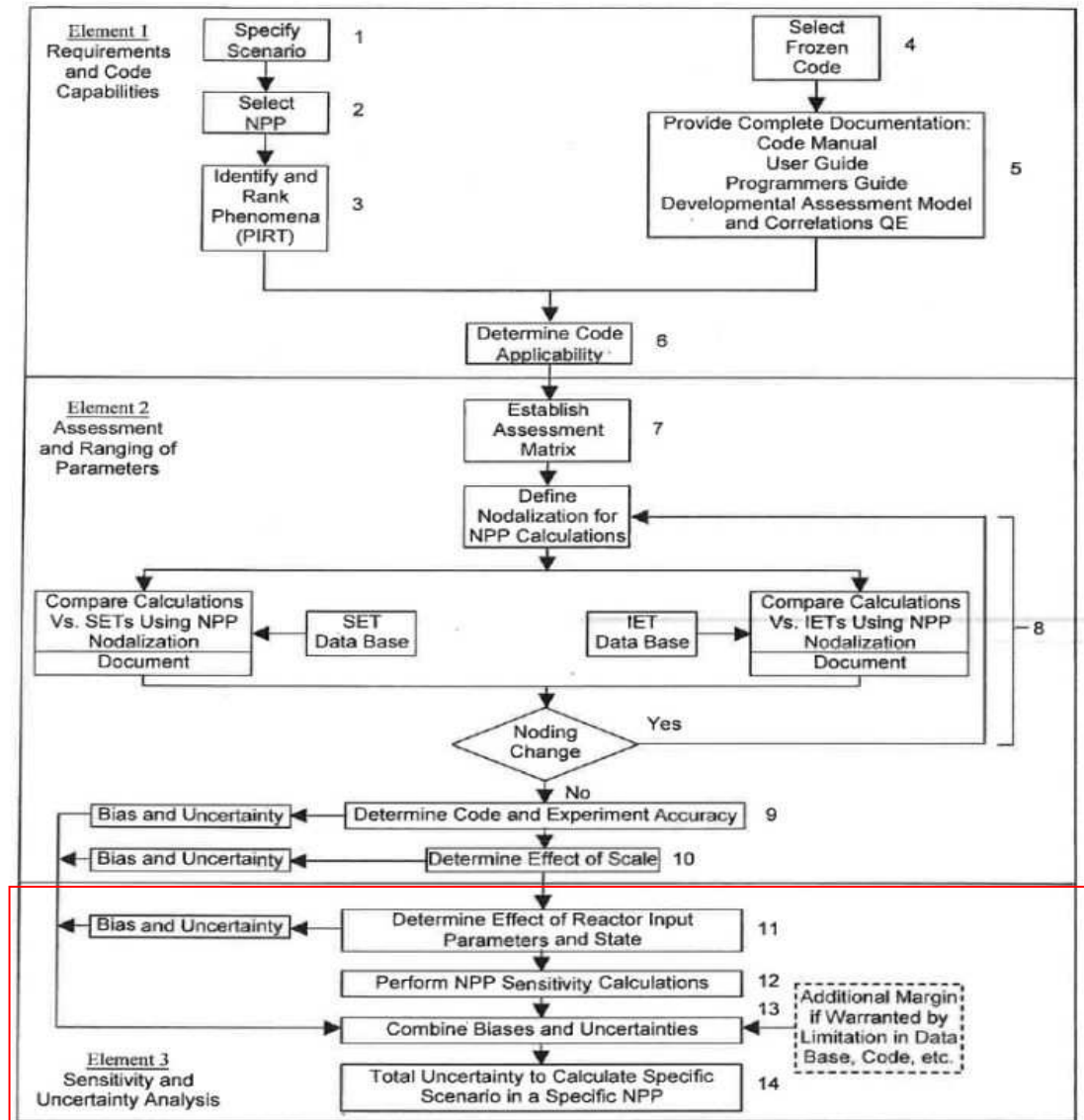
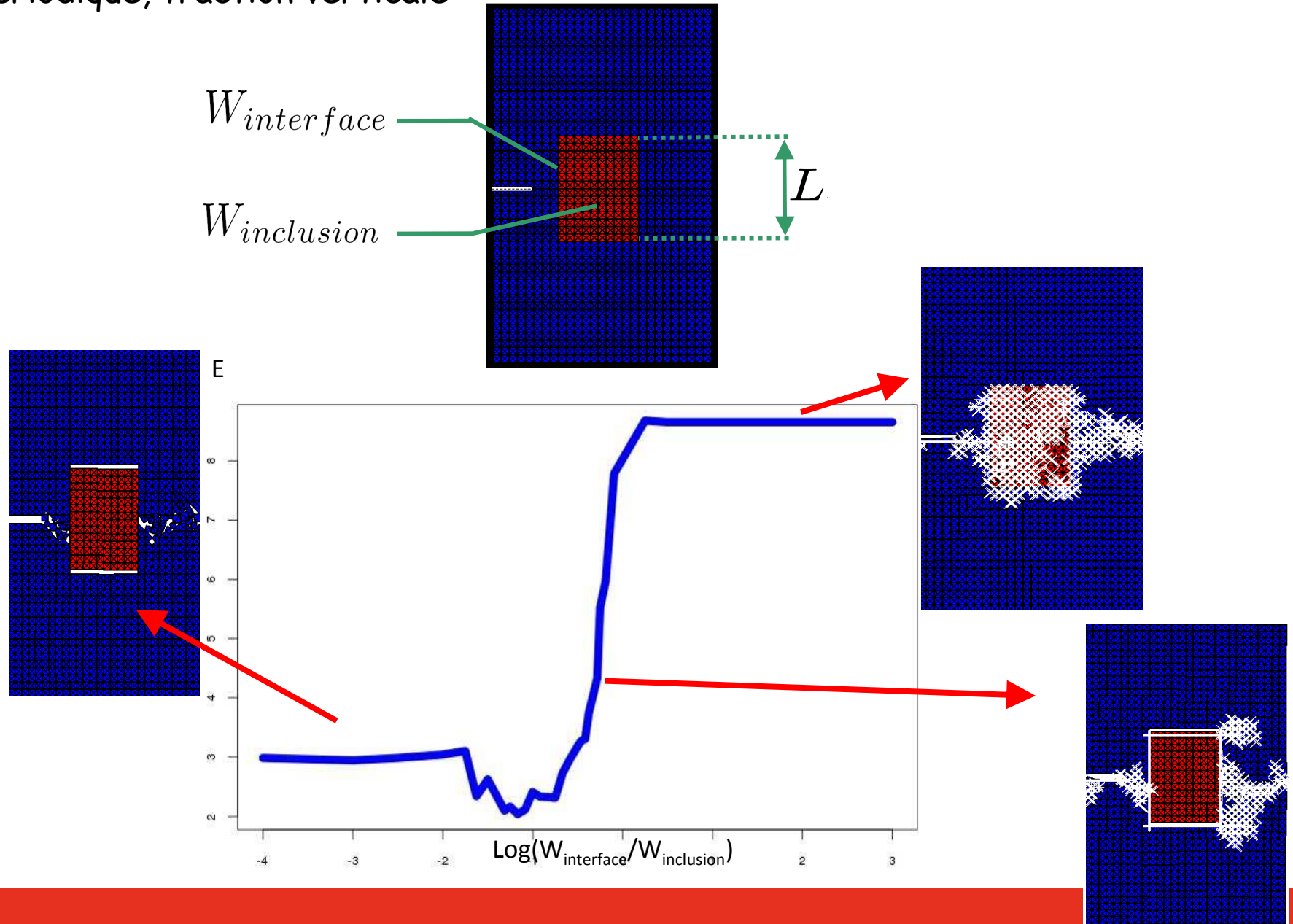


Figure 2.1: Code Scaling, Applicability, and Uncertainty Methodology Flow Chart

Ex (mécanique de la fissuration): simulation à l'aide du logiciel Xper, fissure initiale, maillage périodique, traction verticale



Paramètres et variables d'entrée



Modèle mathématique



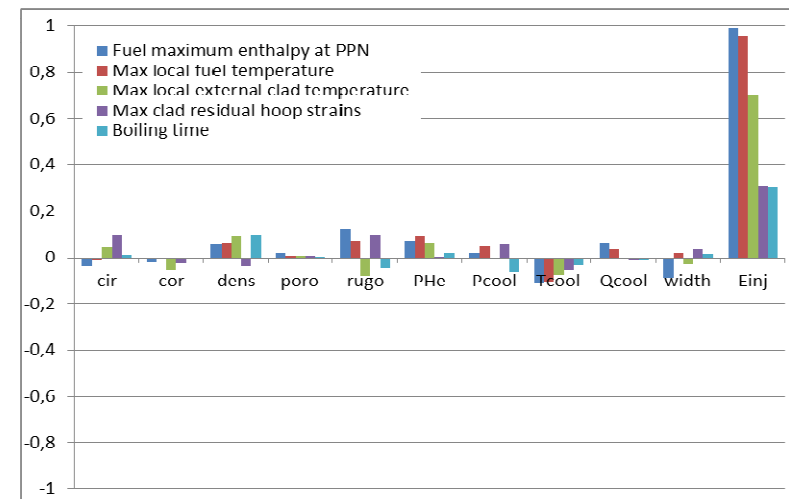
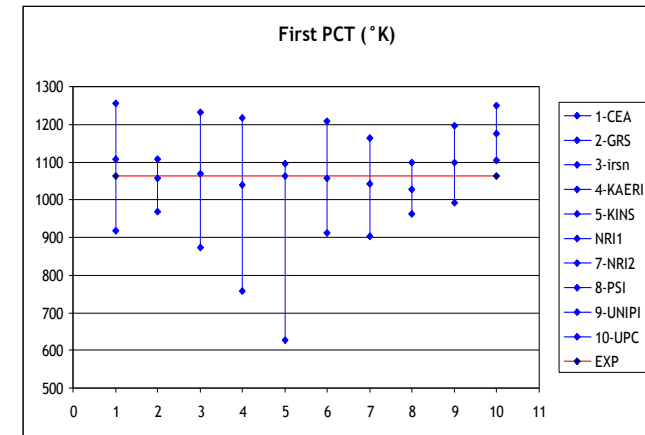
Sorties ou variables d'intérêt



Quantité d'intérêt

Analyse d'incertitude: Quel est l'impact des incertitudes d'entrée sur l'incertitude de sortie?

Analyse de sensibilité: comment la variation d'un modèle peut être distribuée, qualitativement ou quantitativement aux différentes sources de variation?



1) Quelques méthodes pour prendre en compte les incertitudes

- Développement de Taylor à l'ordre 1:

$$Y = f(X_1, X_2)$$

$$Y^* = f(X_1^*, X_2^*), \Delta Y = Y - Y^*, \Delta X_1 = X_1 - X_1^*, \Delta X_2 = X_2 - X_2^*$$

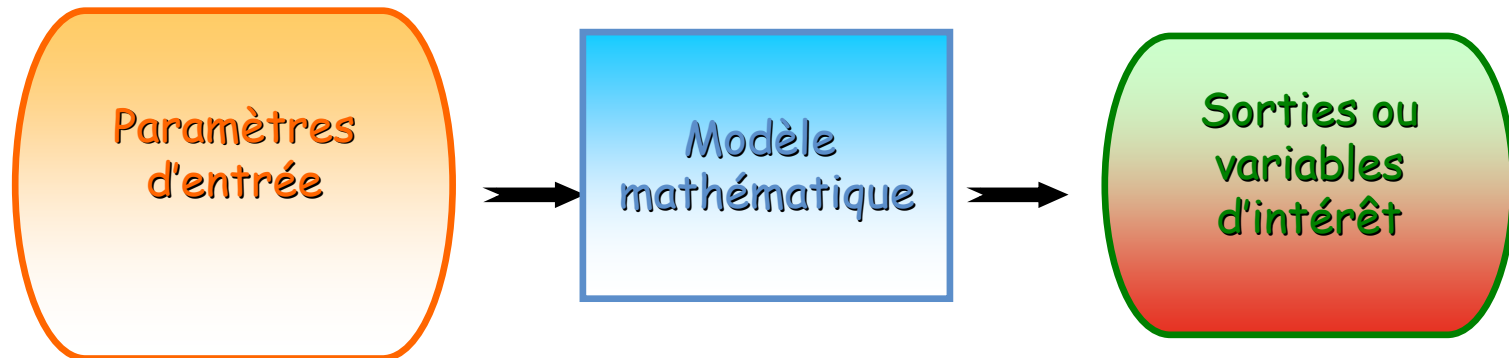
$$\longrightarrow \Delta Y = \frac{\partial f}{\partial x_1} \Delta X_1 + \frac{\partial f}{\partial x_2} \Delta X_2$$

$$\Delta Y = \sqrt{\text{Var}(Y)}, \Delta X_1 = \sqrt{\text{Var}(X_1)}, \Delta X_2 = \sqrt{\text{Var}(X_2)}$$

$$\longrightarrow \Delta Y = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 \Delta X_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \Delta X_2^2}$$

- Hypothèse de linéarité à justifier, Calcul de la dérivée difficile dans le cas où f est un code de calcul complexe

2) Analyse d'incertitude



(Saporta, 1990)

« physique »

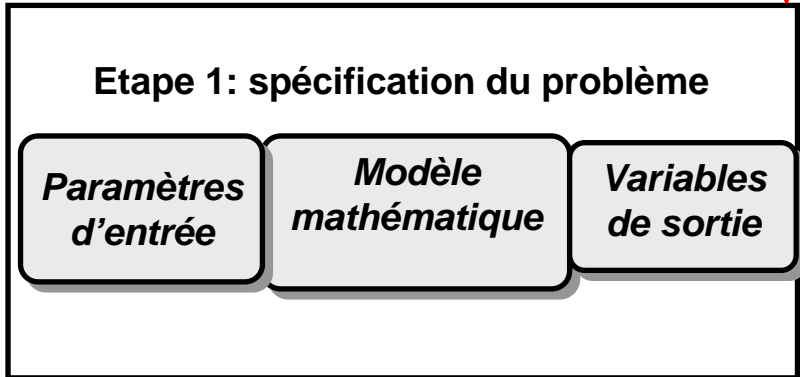
Sources d'incertitudes

X_1, X_2, \dots



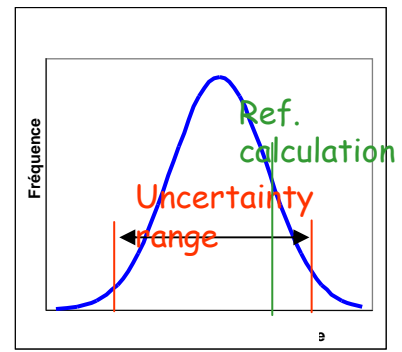
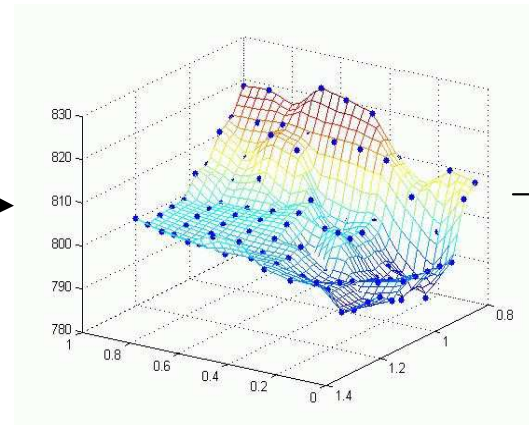
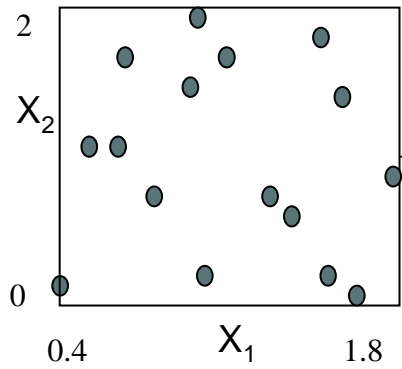
« Traitement de l'information »

Etape 3 : Propagation des incertitudes



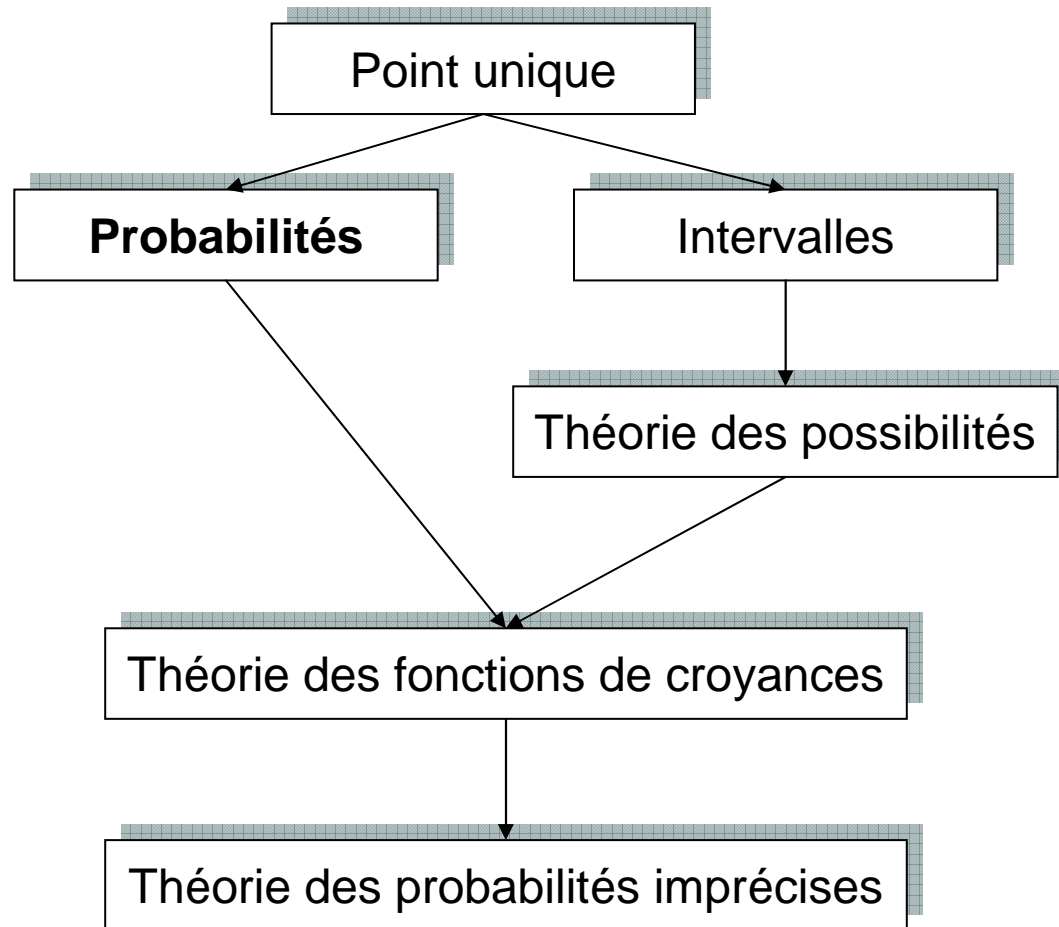
Etape 4 : Analyse statistique des réponses

Paramètre	Nom. Value	Min	Max
X_1	1	0.4	1.8
X_2	1	0	2
...			



Modélisation des incertitudes sources

Plusieurs cadres mathématiques pour la modélisation



Théorie...

+ complexe,

+ générale,

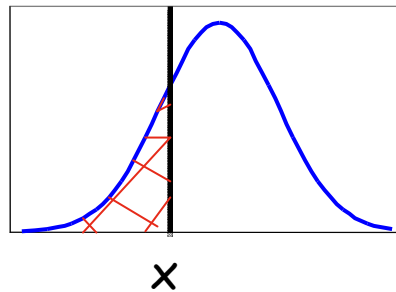
+ flexible,

- facile à mettre en œuvre

Dans les problèmes pratiques, besoin de trouver un compromis entre faisabilité et flexibilité

Rappel: caractérisation d'une variable aléatoire X :

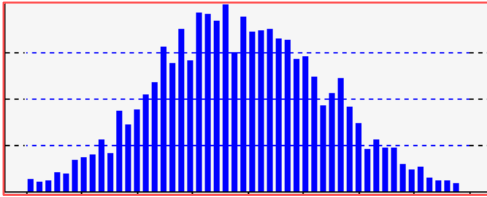
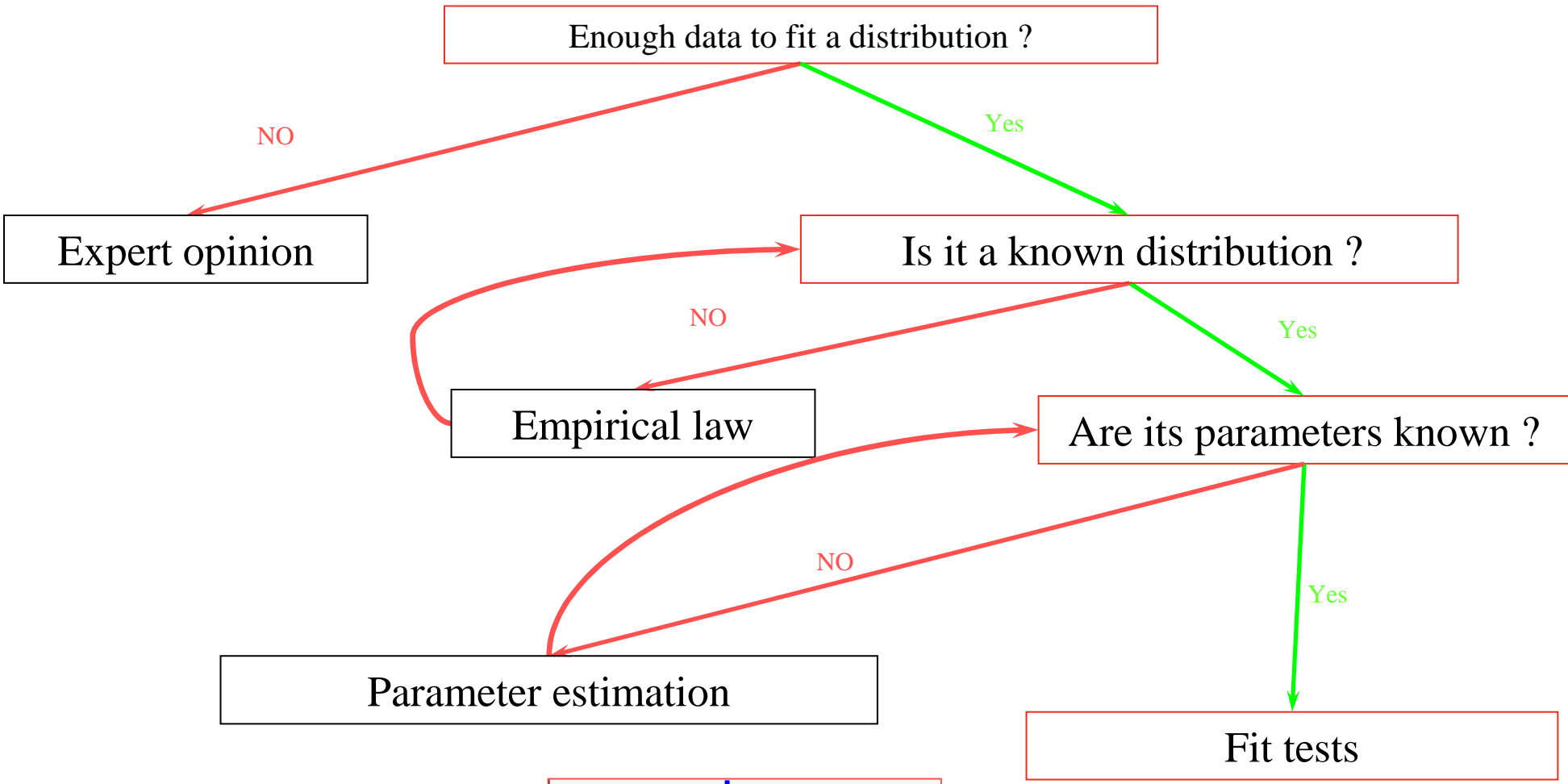
Fonction de répartition $F_X(x) = P(X \leq x)$



Densité de probabilité (variable aléatoire continue): $f_X(x)$

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad f_X(x) = F'_X(x)$$

Avec ou sans données?



Peu de données pour construire empiriquement une distribution








Modélisation par « avis d'experts »

Ex: Recommandations ISO du GUM

JCGM 101:2008

Table 1 — Available information and the PDF assigned on the basis of that information (G.4.1, C.4.2)

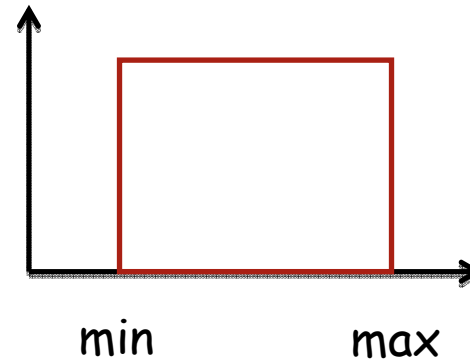
Available information	Assigned PDF and illustration (not to scale)	Subclause
Lower and upper limits a, b	Rectangular: $R(a, b)$ 	G.4.2
Inexact lower and upper limits $a \pm d, b \pm d$	Curvilinear trapezoid: $CTrap(a, b, d)$ 	G.4.3
Sum of two quantities assigned rectangular distributions with lower and upper limits a_1, b_1 and a_2, b_2	Trapezoidal: $Trap(a, b, \beta)$ with $a = a_1 + a_2,$ $b = b_1 + b_2,$ $\beta = (b_1 - a_1) - (b_2 - a_2) / (b - a)$ 	G.4.4
Sum of two quantities assigned rectangular distributions with lower and upper limits a_1, b_1 and a_2, b_2 and the same semi-width ($b_1 - a_1 = b_2 - a_2$)	Triangular: $T(a, b)$ with $a = a_1 + a_2, b = b_1 + b_2$ 	G.4.5
Sinusoidal cycling between lower and upper limits a, b	Arc sine (U-shaped): $U(a, b)$ 	G.4.6
	^ ...	

Valeurs min et max : loi uniforme

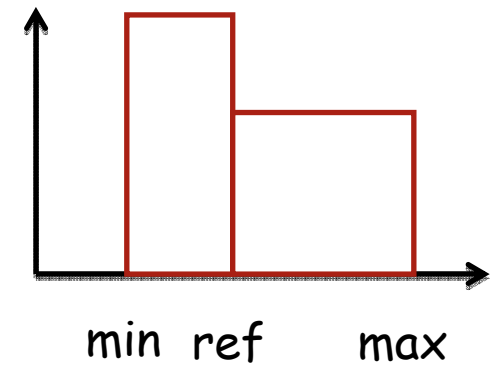
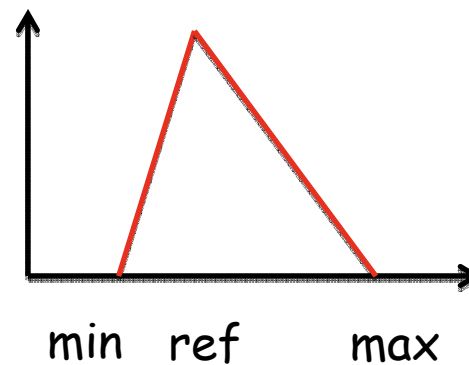
Valeurs min et max + Valeur à l'intérieur : loi triangulaire, histogramme

Moyenne et variance: loi normale

Valeurs min et max : loi uniforme



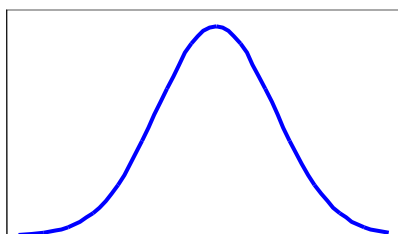
Valeurs min et max + Valeur à l'intérieur : loi triangulaire, histogramme



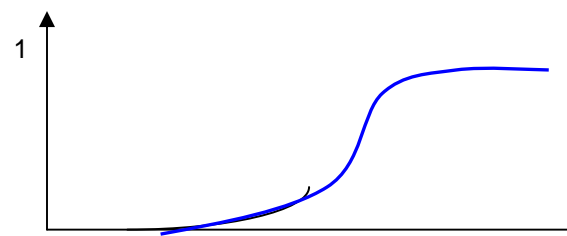
Moyenne et variance: loi normale

Comment tirer une valeur à partir d'une loi de probabilité?

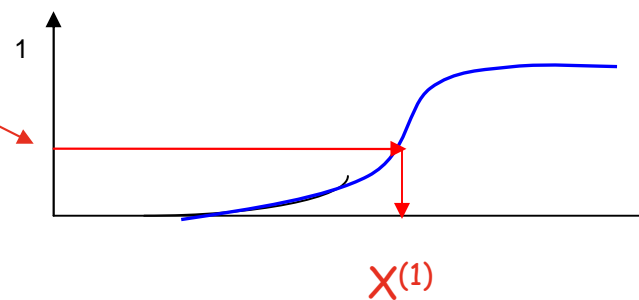
Densité de probabilité



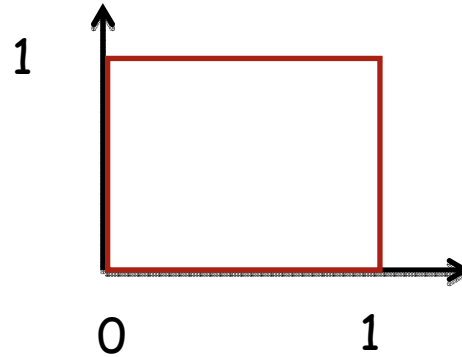
Fonction de répartition $F_X(x) = P(X \leq x)$



Tirer une valeur entre 0 et 1



Exo1: Fonction de répartition de la loi uniforme sur $[0,1]$?



$$F_X(x) = P(X \leq x) \longrightarrow F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Exo2: Si X est une variable aléatoire quelconque de fonction de répartition F , montrer que $F(X)$ suit une loi uniforme sur $[0,1]$.

$$Y=F(X)$$

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(F(X) \leq x) \\ &= P(X \leq F^{-1}(x)) \\ &= \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases} \end{aligned}$$

Pour tirer une valeur:

1) Tirer une valeur à partir de la loi uniforme sur $[0,1]$ $\longrightarrow u_i$

2) « Inverser » la fonction de répartition de X $\longrightarrow x_i = F^{-1}(u_i)$

Exo3: Calculer la loi de $1/X$ où X suit une loi uniforme sur $[a,b]$, $a>0$.

$$X \infty U([a,b])$$

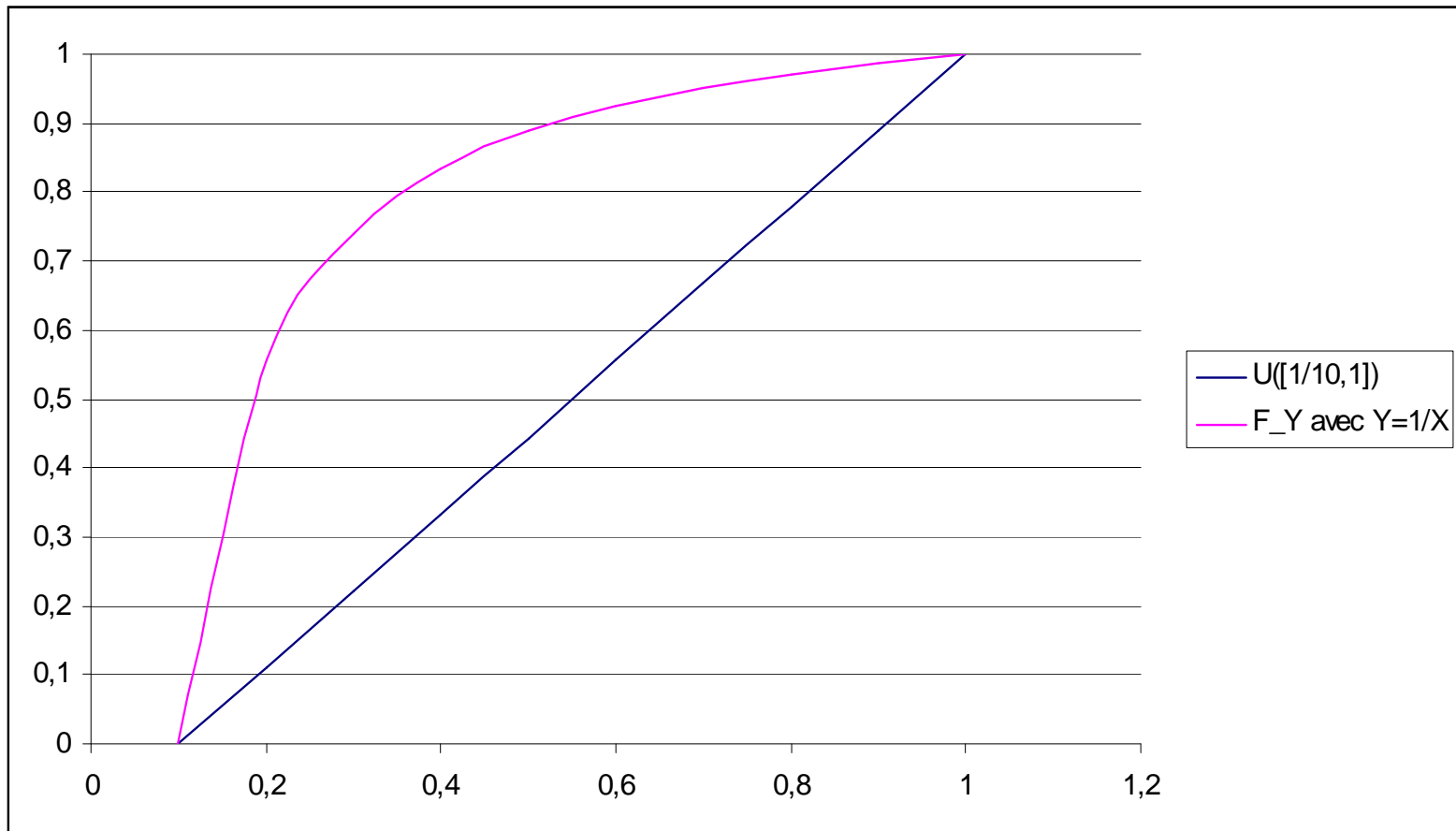
$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x)$$

$$\begin{aligned} Y = \frac{1}{X} &\longrightarrow F_Y(x) = P\left(\frac{1}{X} \leq x\right) \\ &= P\left(X \geq \frac{1}{x}\right) \\ &= 1 - P\left(X < \frac{1}{x}\right) \\ &= 1 - F_X\left(\frac{1}{x}\right) \\ &= \begin{cases} 0 & x < \frac{1}{b} \\ \frac{b - 1/x}{b - a} & \frac{1}{b} \leq x \leq \frac{1}{a} \\ 1 & x > \frac{1}{a} \end{cases} \end{aligned}$$

$$Y \infty U\left(\left[\frac{1}{b}, \frac{1}{a}\right]\right)$$

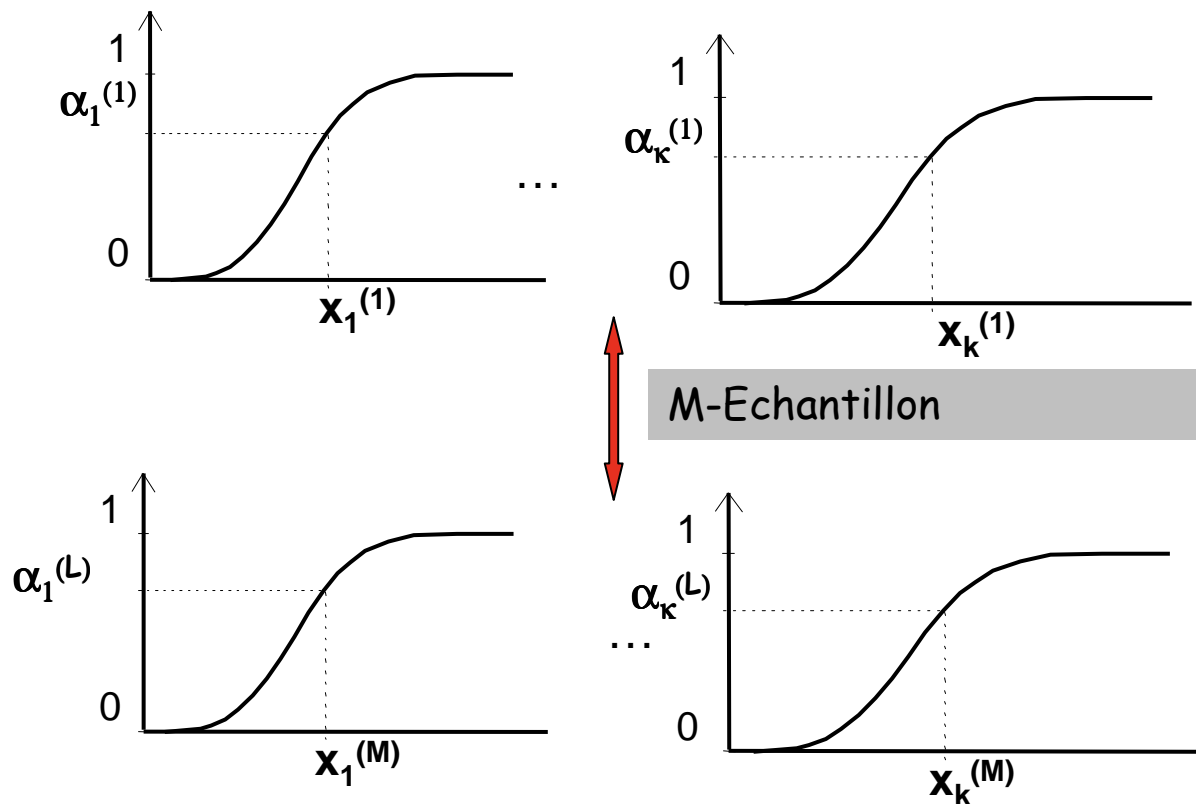
$$F_Y(x) = \begin{cases} 0 & x < \frac{1}{b} \\ \frac{x - \frac{1}{b}}{\frac{1}{a} - \frac{1}{b}} & \frac{1}{b} \leq x \leq \frac{1}{a} \\ 1 & x > \frac{1}{a} \end{cases}$$

Ex: $a=1, b=10$



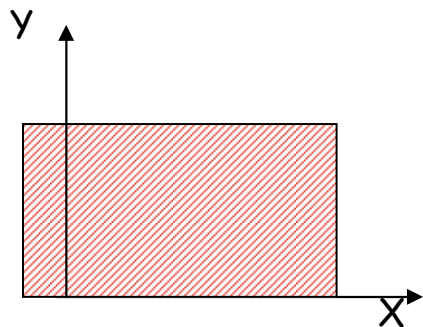
Construction de la matrice échantillon

- Des valeurs des paramètres incertains sont tirées à partir des lois de probabilité associées

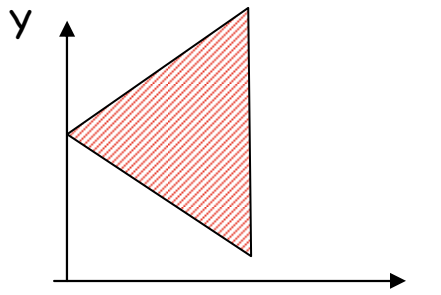


- Prise en compte des dépendances (Iman&Conover, Copules,...)

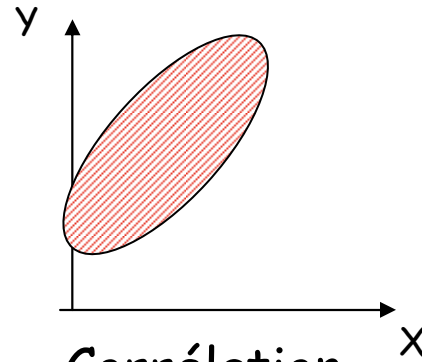
Comment quantifier la corrélation entre 2 variables?



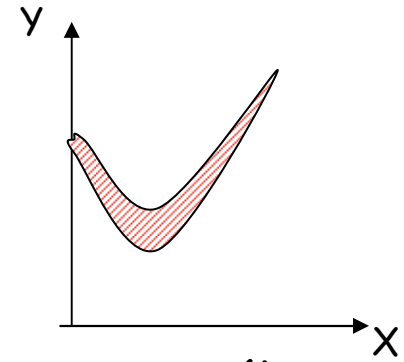
Absence de liaison



Absence de liaison en moyenne mais pas en dispersion



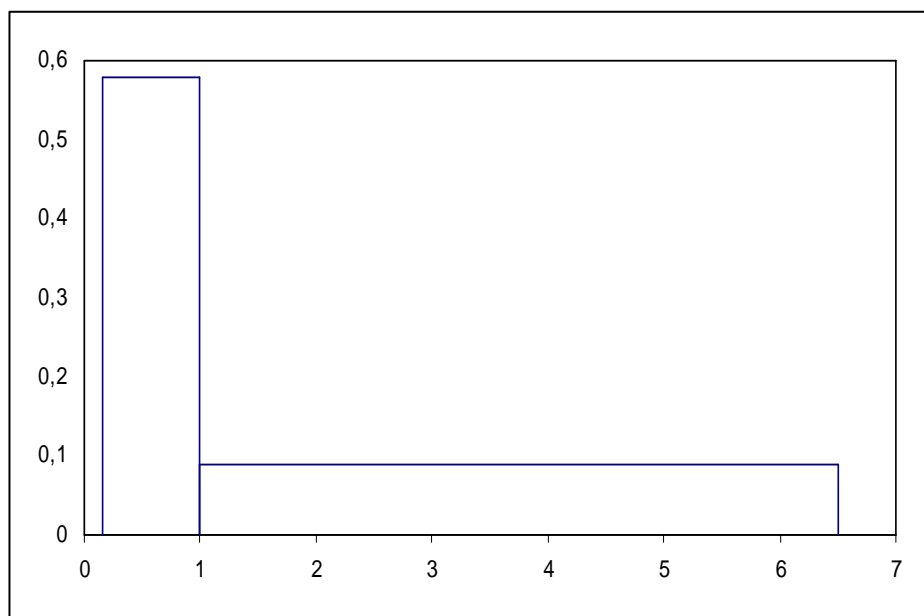
Corrélation linéaire positive



Corrélation non linéaire

Impact du choix des lois sur la construction de la matrice échantillon

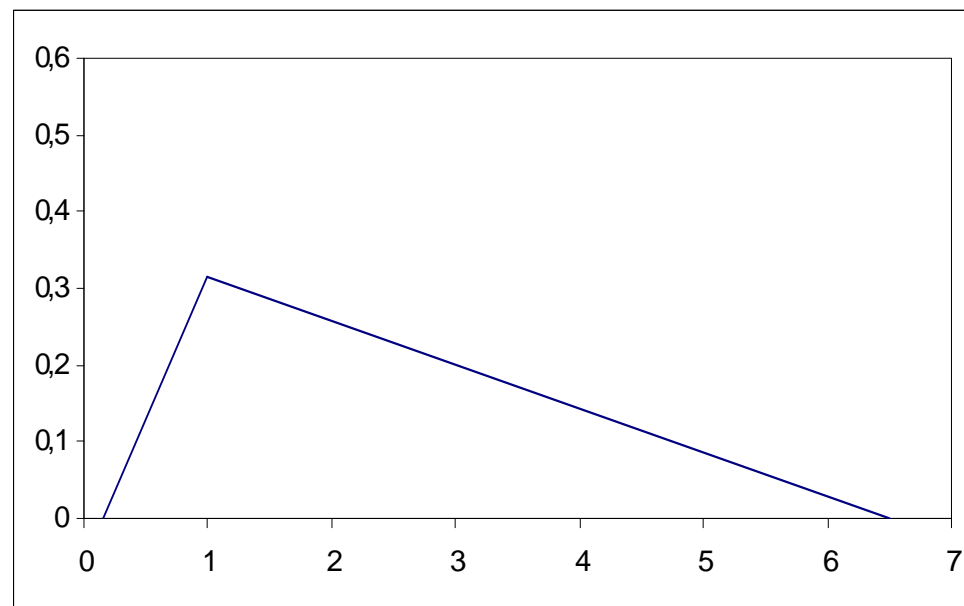
Ex: paramètre P1 qui varie entre 0.15 et 6.5 avec 1 comme valeur de référence



Symétrie:

50% dans [0,15;1]

50% dans [1;6,5]



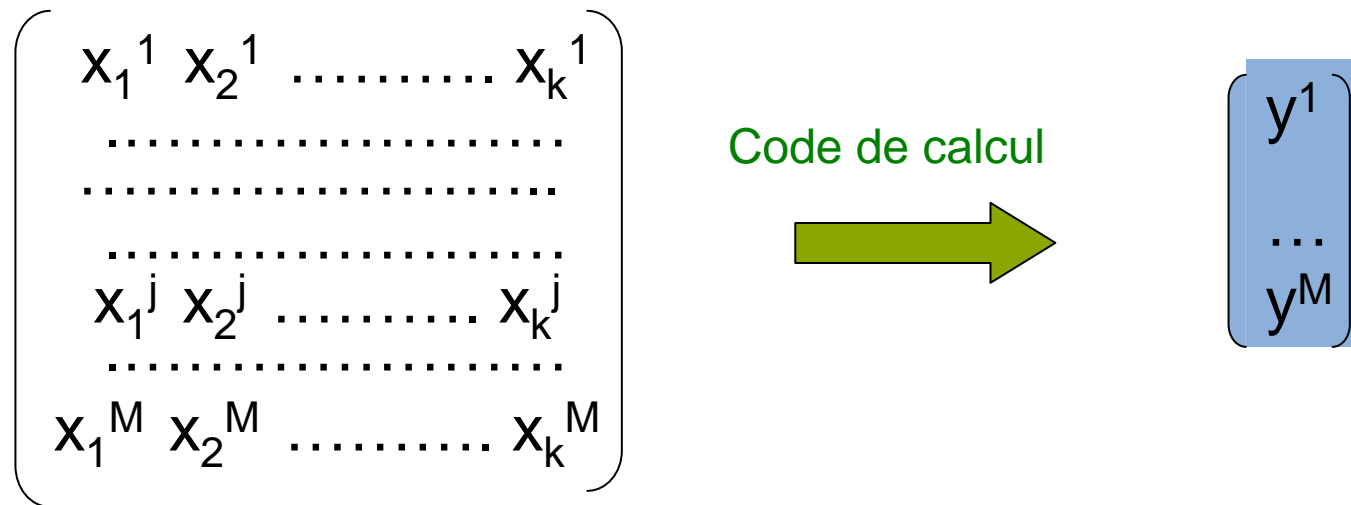
Pas de symétrie:

13% dans [0,15;1]

87% dans [1;6,5]

Propagation des incertitudes

Lancement des simulations à l'aide du code de calcul (propagation de type Monte Carlo)



Analyse statistique

Une quantité incertaine dont on a observé plusieurs réalisations

y^1
...
 y^M

Quantités statistiques usuelles et leur estimateur:

Moyenne : $\bar{Y} = \frac{\sum_{i=1}^M Y_i}{M}$

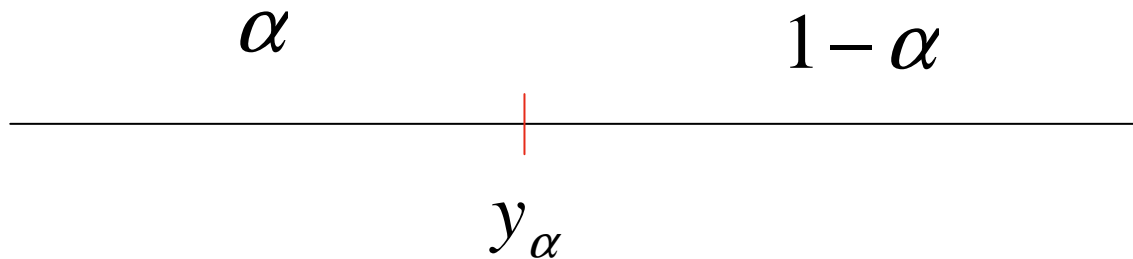
Variance : $\sigma^2 = \frac{\sum_{i=1}^M (Y_i - \bar{Y})^2}{M - 1}$

Fonction de répartition : $F_Y(x) = \frac{\sum_{i=1}^M 1_{[y_i, +\infty[}(x)}{M}$



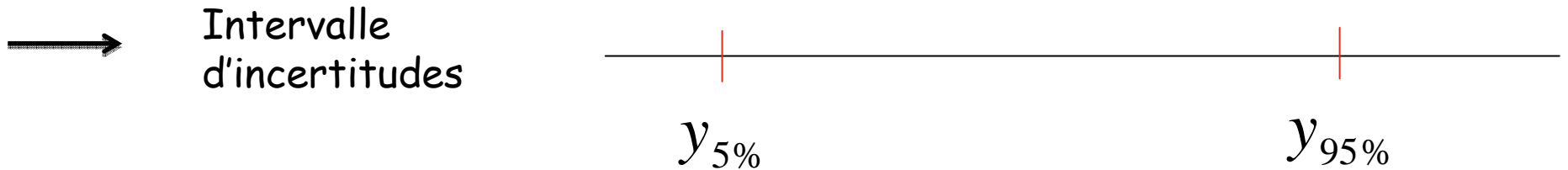
Percentiles et statistiques d'ordre

$$y_\alpha \text{ tel que } F_Y(y_\alpha) = \alpha$$



α grand ($\alpha=0,95$): permet de trouver un majorant de Y avec une forte probabilité

α petit ($\alpha=0,05$): permet de trouver un minorant de Y avec une forte probabilité



Statistiques d'ordre:

Classer par ordre croissant les valeurs de l'échantillon

$$y^{(1)} \leq \dots \leq y^{(M)}$$

 2 types d'information

1) $y_\alpha \approx y^{(\alpha M)}$

N percentile	100	200	300	400	500	600	700	800	900	1000
0,9	90	180	270	360	450	540	630	720	810	900
0,95	95	190	285	380	475	570	665	760	855	950
0,975	98	195	293	390	488	585	683	780	878	975
0,99	99	198	297	396	495	594	693	792	891	990

$$2) \quad P(y^{(k)} \leq y_\alpha) = F_{\beta(k, M-k+1)}(\alpha)$$

Densité de la loi Béta :

$$\beta_{(k, M-k+1)}(x) = M! / [(k-1)!(M-k)!] x^{k-1}(1-x)^{M-k}$$

Fonction de répartition de la loi Béta :

$$F_{\beta(k, M-k+1)}(x) = \sum_{i=k}^M C_M^i x^i (1-x)^{M-i}$$



Loi de Bernoulli de paramètre p : loi d'une variable aléatoire X ne pouvant prendre que 2 valeurs 1 ou 0 avec les probabilités p et $1-p$

Loi binomiale $B(n,p)$: somme de n variables de Bernoulli indépendantes et de même paramètre p

$$Y \propto B(n, p) \qquad Y = \sum_{i=1}^n X_i \quad \text{où} \quad X_i \propto B(p)$$

$$\longrightarrow P(Y \geq k) = \sum_{i=k}^n C_n^i p^i (1-p)^{n-i}$$

Exo4: Montrer que pour une variable aléatoire **quelconque** dont on observe M réalisations

$$P(X^{(k)} \leq X_\alpha) = \sum_{i=k}^M C_M^i \alpha^i (1-\alpha)^{M-i}$$

Reformuler la probabilité à calculer comme une probabilité faisant intervenir une variable binomiale

$$Y = \begin{cases} 1 & X \leq X_\alpha \\ 0 & X > X_\alpha \end{cases} \infty B(\alpha) \quad Z = \sum_{i=1}^M Y_i \infty B(M, \alpha)$$

$$P(X^{(k)} \leq X_\alpha) = P(Z \geq k) = \sum_{i=k}^M C_M^i \alpha^i (1-\alpha)^{M-i}$$

Application 1: trouver la taille minimale M pour obtenir un majorant d'un percentile avec un niveau de confiance γ $P(Y_\alpha \leq Y^{(M)}) = \gamma$

$$P(X^{(k)} \leq X_\alpha) = \sum_{i=k}^M C_M^i \alpha^i (1-\alpha)^{M-i}$$

$$P(X^{(M)} > X_\alpha) = 1 - \sum_{i=M}^M C_M^i \alpha^i (1-\alpha)^{M-i} = 1 - \alpha^M$$

$$1 - \alpha^M \geq \gamma \Leftrightarrow M \geq \frac{\ln(1-\gamma)}{\ln(\alpha)}$$

“Formule de Wilks”

Table : Taille minimale de l'échantillon pour trouver un majorant du percentile α avec un niveau de confiance γ

γ	α	0.80	0.90	0.95	0.99
	β				
0.80		8	16	32	161
0.90		11	22	45	230
0.95		14	29	59	299
0.99		21	44	90	459

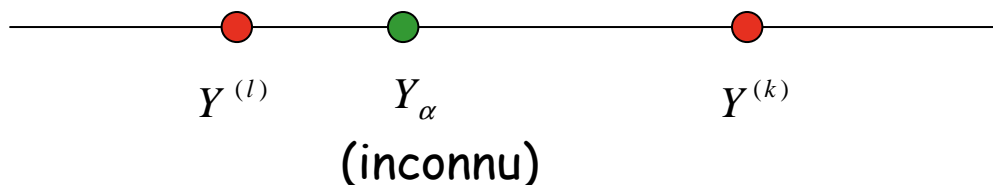
Application 2: pour une taille de vecteur échantillon fixée, trouver un minorant ou/et un majorant d'un percentile

M fixé, trouver k ou/et l tel que

$$P(Y^{(k)} > y_\alpha) = \gamma$$

$$P(Y^{(l)} \leq y_\alpha) = \gamma$$

Ex: γ grand, estimation de l'incertitude numérique



Exo5 : taille fixée à 100, trouver l'indice k tel que

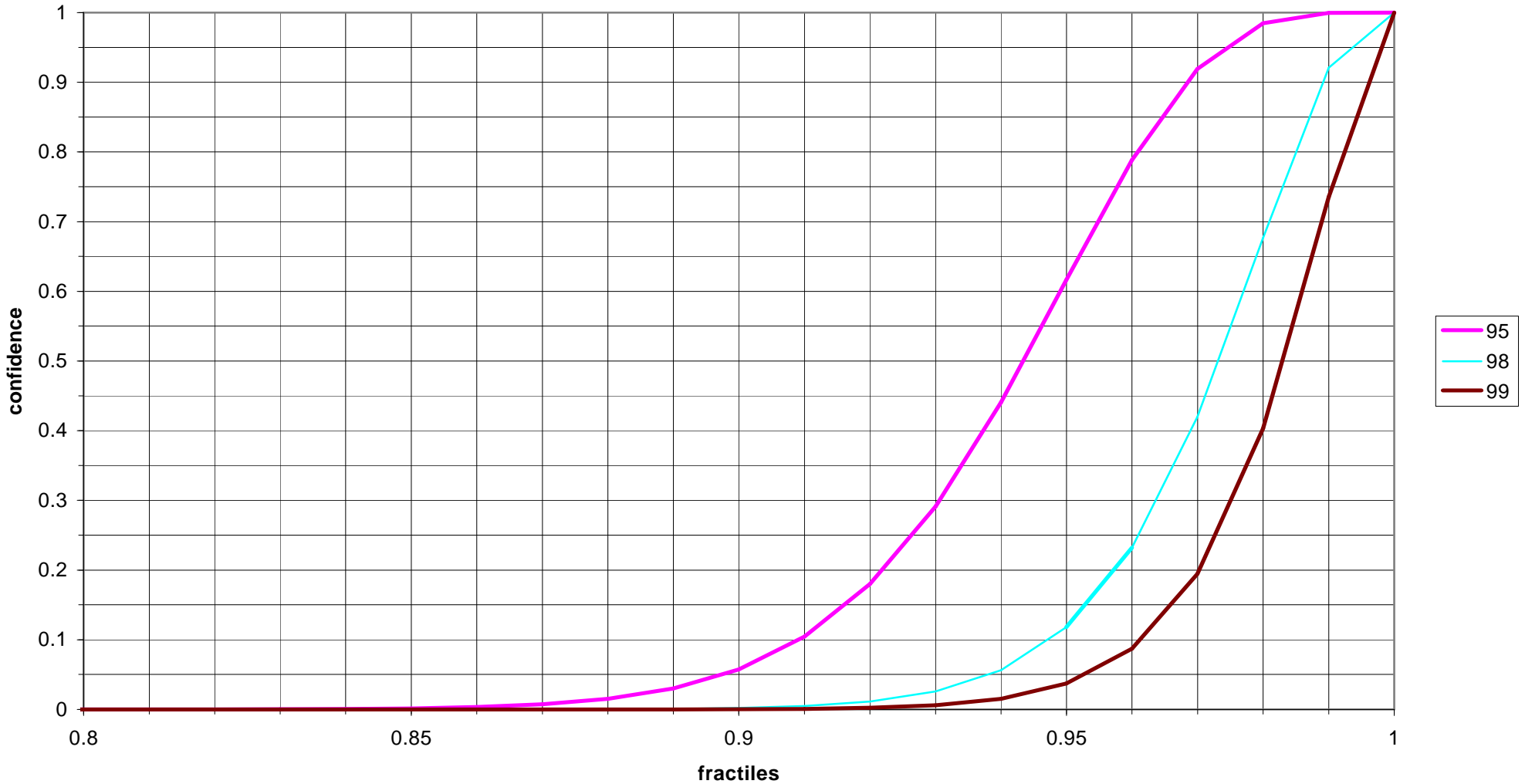
$$P(Y^{(k)} > y_{95\%}) \approx 96\%$$

$$F_{\beta(95, 100-95+1)}(\alpha)$$

$$F_{\beta(98, 100-98+1)}(\alpha)$$

$$F_{\beta(99, 100-95+1)}(\alpha)$$

beta



Conclusion sur l'analyse d'incertitude probabiliste

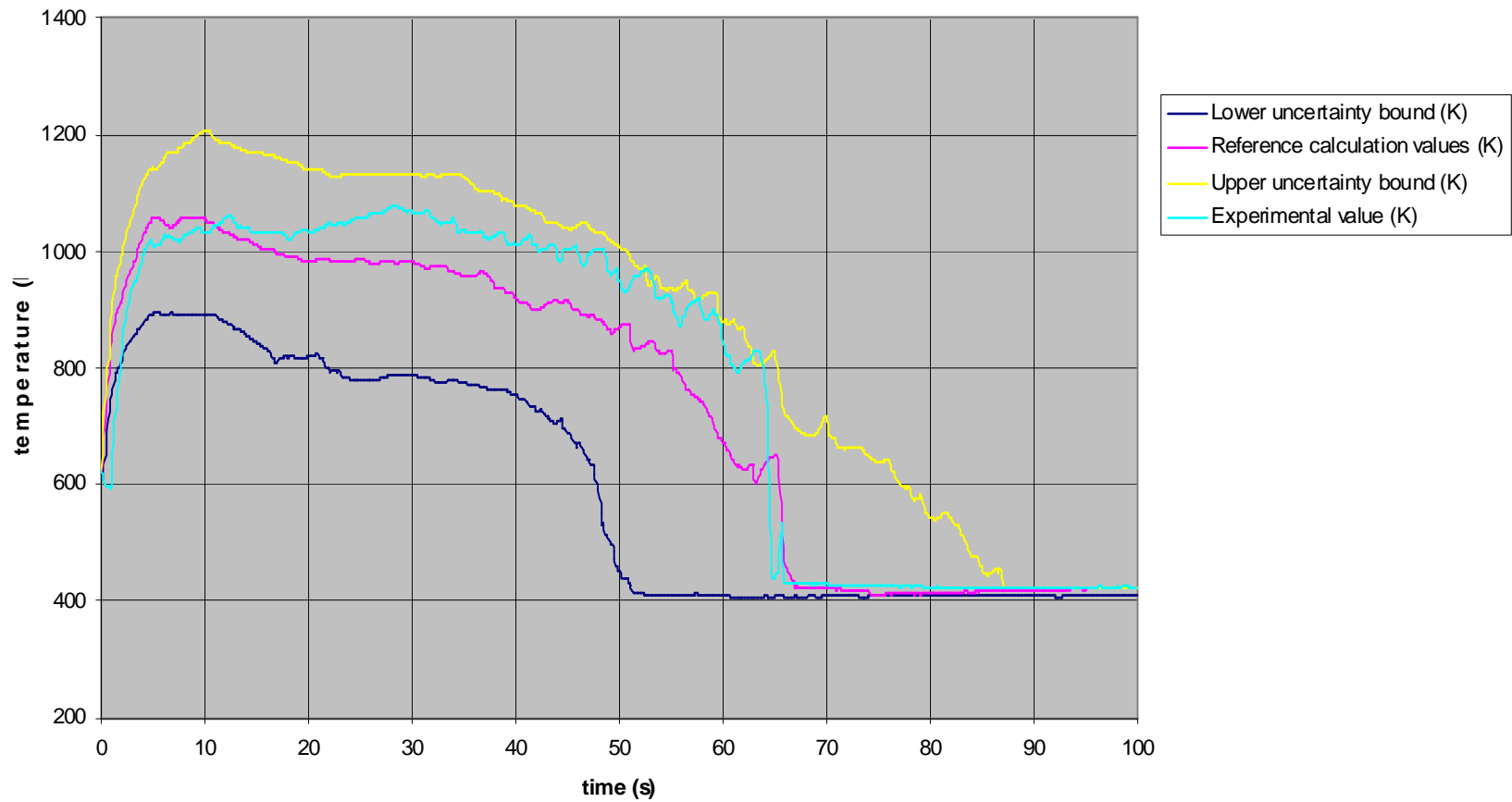
- Mise en oeuvre aisée (Monte-Carlo), estimation directe des percentiles à partir des statistiques d'ordre
- La modélisation des incertitudes (choix lois de probabilité et structure de dépendance) est l'étape clé de l'analyse, l'état de connaissance n'est souvent que partiel
- 2 Exemples: Programme BEMUSE, GRNC

BEMUSE: Application des méthodologies de prise en compte des incertitudes dans des codes de thermohydrauliques simulant les phénomènes complexes se déroulant au sein d'un réacteur nucléaire

GRNC: Risque de leucémie imputable aux activités nucléaire de la Hague

BEMUSE

Exemple de résultats fournis par l'analyse d'incertitude

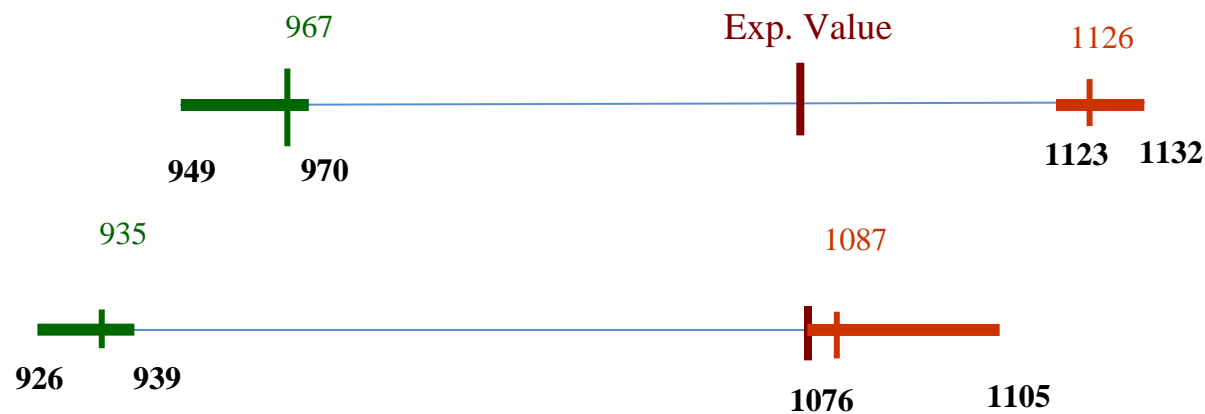


Choix des lois de probabilité:

Estimation des percentiles à 5% et 95% associés au pic de température
(Baccou, Chojnacki, 2007)

Parameter	Nom. Value	Variation Range
Liquid-wall friction	1	[0.8;1.9]
Vapour-wall heat transfer (forced convection regime)	1	[0.5;2]
Film boiling (Berenson/Bryce) standard model	1	[0.15;6.5]

Histogram/Uniform pdfs

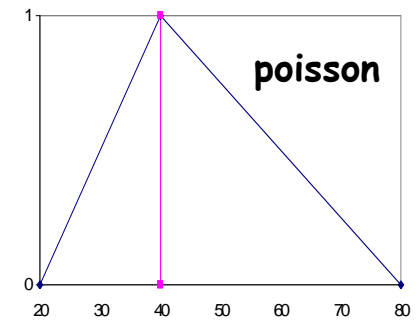
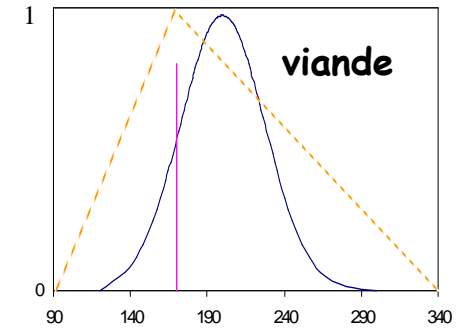
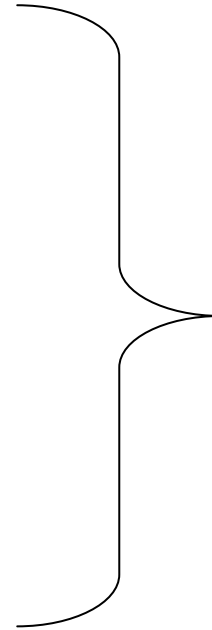
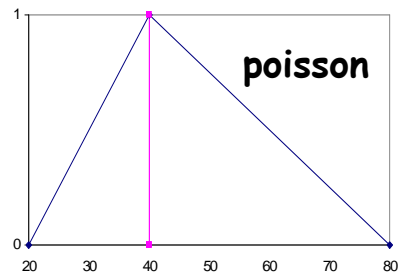
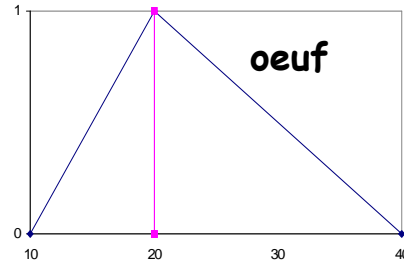
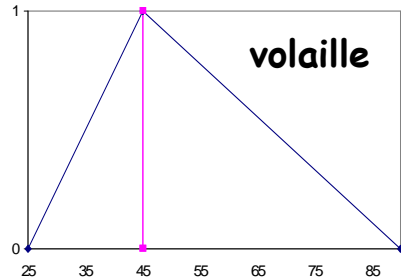
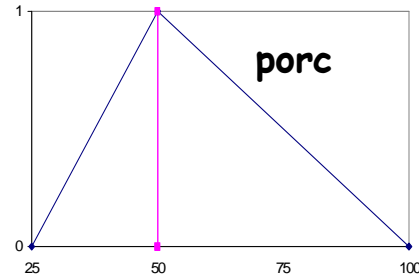
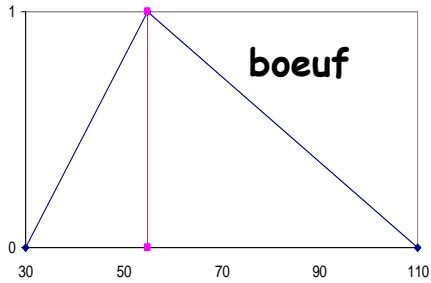


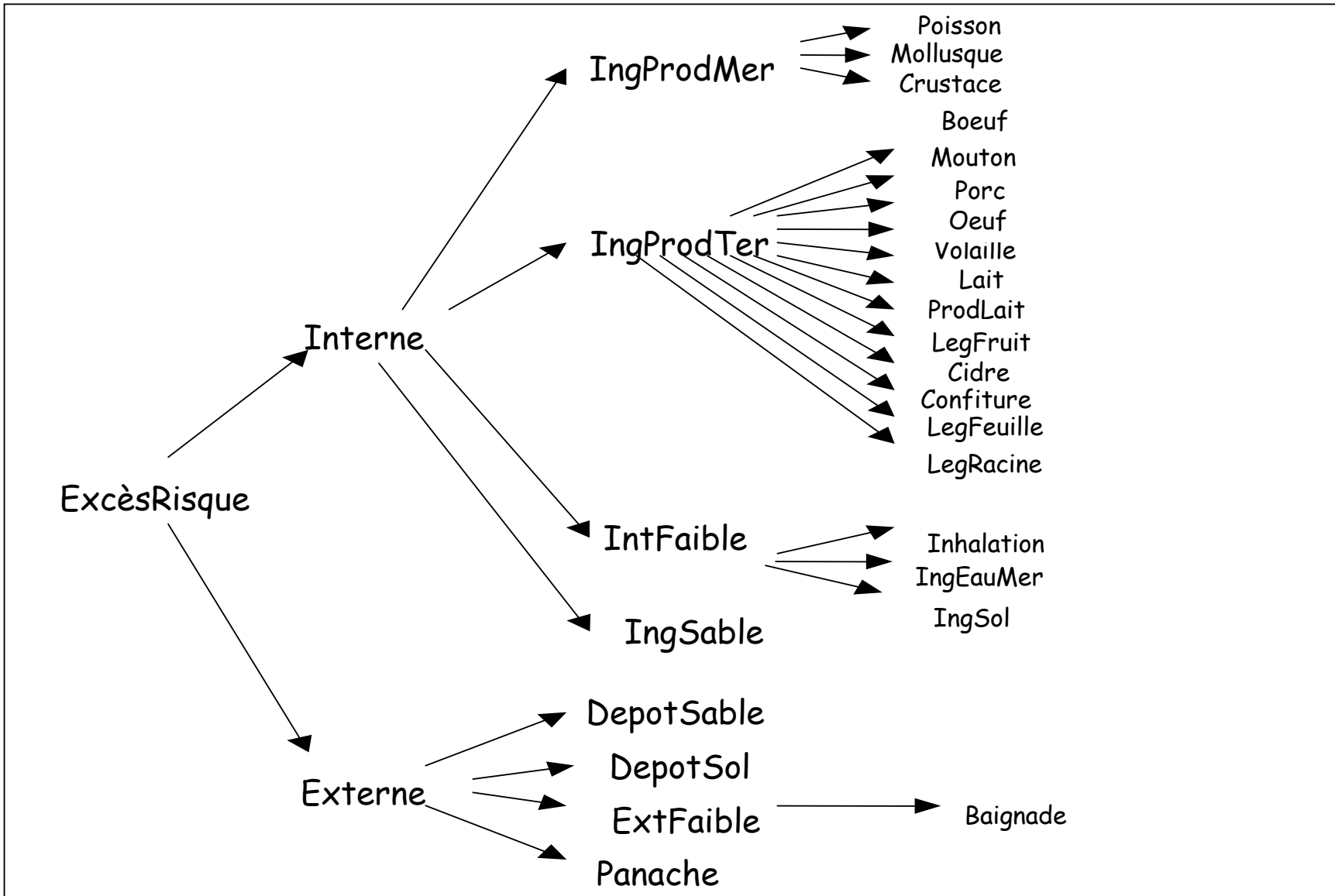
(Chojnacki, Baccou, Destercke, 2010)

GRNC

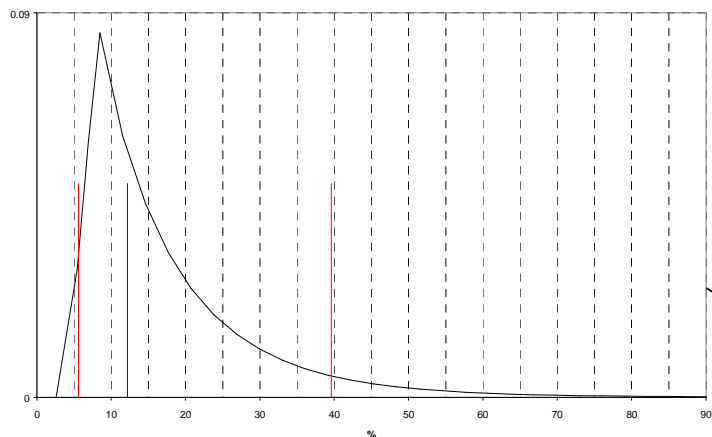
Structure de dépendance:

Exemple: quantité de viande/poisson consommée

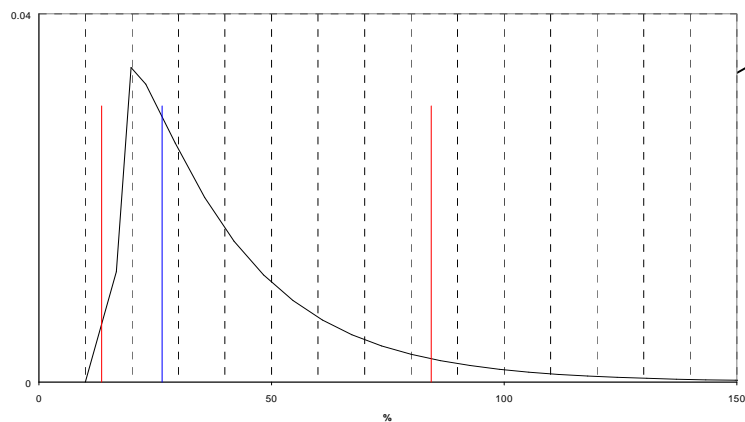




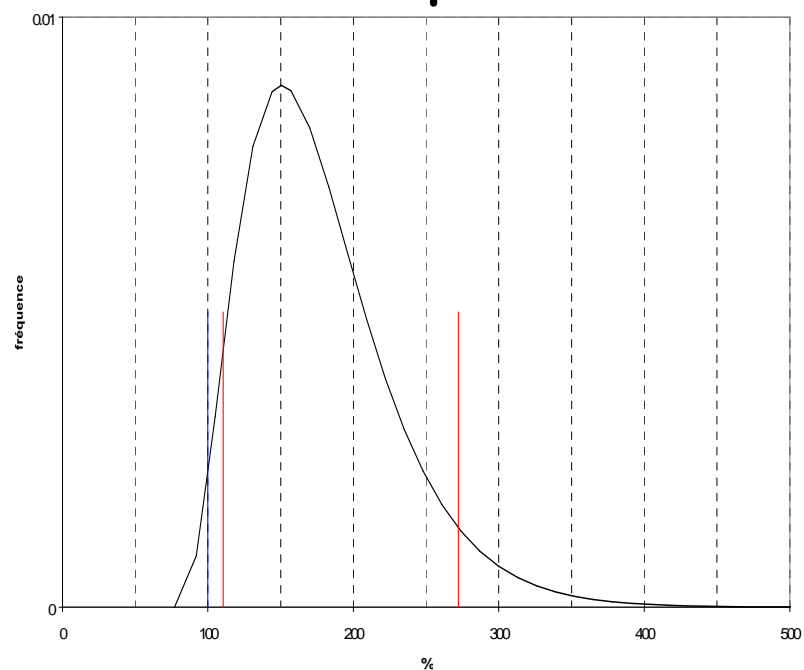
Risque par ingestion de crustacés



Risque par ingestion de poissons



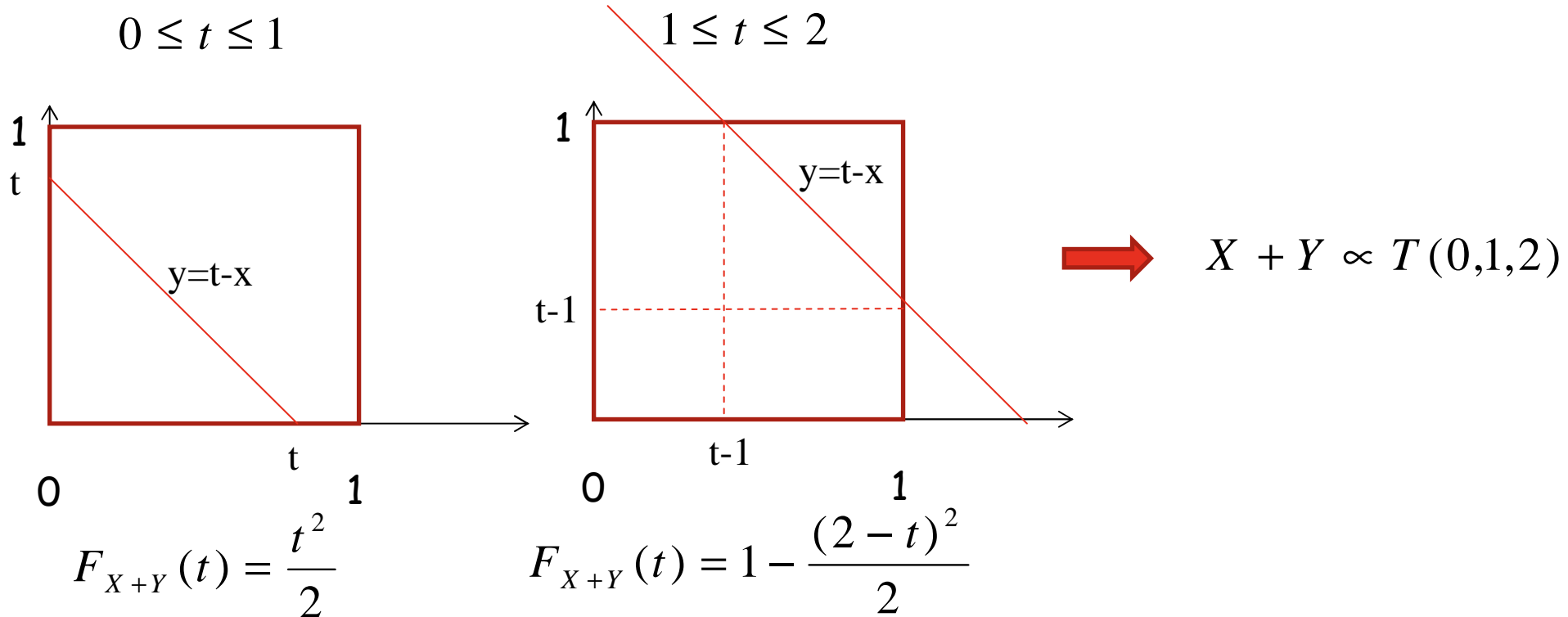
Excès de risque total



En agrégeant les composants de risque, les incertitudes se compensent artificiellement du fait des hypothèses d'indépendance

Exo5: Soient X et Y , 2 variables aléatoires suivant une loi uniforme sur $[0,1]$. Quelle est la loi de $X+Y$ en supposant X et Y indépendants?

$$F_{X+Y}(t) = P(X + Y \leq t) = P(Y \leq t - X)$$



3) Analyse de sensibilité

(Saltelli, Chan, Scott, 2000)

(Iooss, Lemaître, 2015)

Catégories de méthodes

- Criblage: identifier les paramètres d'entrée influents dans un ensemble avec beaucoup de paramètres
- Analyse de sensibilité locale: estimer l'influence locale de chaque paramètre (petite perturbation) sur les sorties du modèle
 - Dérivées partielles
- Analyse de sensibilité globale: estimer l'influence de chaque paramètre sur les sorties du modèle en prenant en compte tout l'ensemble de variation du paramètre (**Régression**, Sobol, ...)

Régression multi-linéaire

(Box, Draper, 1987)

$$Y = \begin{pmatrix} y^1 \\ \dots \\ y^M \end{pmatrix} = \begin{pmatrix} x_1^1 \dots x_k^1 \\ \dots \\ x_1^M \dots x_k^M \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_M \end{pmatrix} = X\alpha + \varepsilon$$

$$Y^* = X\alpha$$

Moindres Carrés $\longrightarrow \alpha = ({}^tXX)^{-1} {}^tXY$

Construction de la matrice échantillon

- Echantillonnage aléatoire issu de simulations Monte-Carlo
- Plans d'expériences (*Benoist, Tourbier, Germain-Tourbier, 1994*)

But: étudier la relation R existant entre les variables X d'un système et les réponses Y , d'en évaluer les paramètres avec un maximum de précision et en un minimum d'essais

- Ecriture d'un modèle formel : $Y = X_1 + X_2 + X_3 + X_1X_2 + X_1X_3$ (résulte d'une analyse physique: effets à prendre en compte?)
- Construction de la matrice échantillon

(Suard, Hostikka, Baccou, 2013)

Mesures de sensibilité: coefficients de corrélation

- Coefficient de corrélation linéaire ou de Bravais-Pearson (SCC):

Degré de dépendance linéaire entre la réponse et chacun des régresseurs (paramètres d'entrée) pris séparément

$$\rho_i = \frac{Cov(Y, X_i)}{\sigma_{X_i} \sigma_Y} \quad -1 \leq \rho_i \leq 1$$

$$Y = aX_i + b \quad \longrightarrow \quad \rho_i = \pm 1$$

- Coefficient de corrélation partielle (PCC):

Degré de dépendance linéaire entre un régresseur et la réponse à laquelle on a soustrait la partie expliquée par tous les autres régresseurs

$$\rho_i = \frac{Cov(Y - \hat{Y}_{\sim i}, X_i)}{\sigma_{X_i} \sigma_{Y - \hat{Y}_{\sim i}}}$$

$\hat{Y}_{\sim i}$: valeurs données par le modèle de régression dans lequel on a supprimé la colonne i de la matrice X



Mesures de dépendance linéaire



Si indépendance entre Y et X_i alors $\rho=0$ mais faux dans l'autre sens

$$\rho_i = \frac{\text{Cov}(Y, X_i)}{\sigma_{X_i} \sigma_Y}$$

Exo6: on considère 2 variables aléatoires X, Y telles que $Y=aX+b$.
Calculer le coefficient de corrélation linéaire

$$\begin{aligned}\text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E((X - E(X))(aX - aE(X))) \\ &= a E((X - E(X))^2) \\ &= a \sigma_X^2\end{aligned}$$

$$\begin{aligned}\sigma_Y^2 &= E((Y - E(Y))^2) \\ &= a^2 E((X - E(X))^2) \\ &= a^2 \sigma_X^2\end{aligned}$$

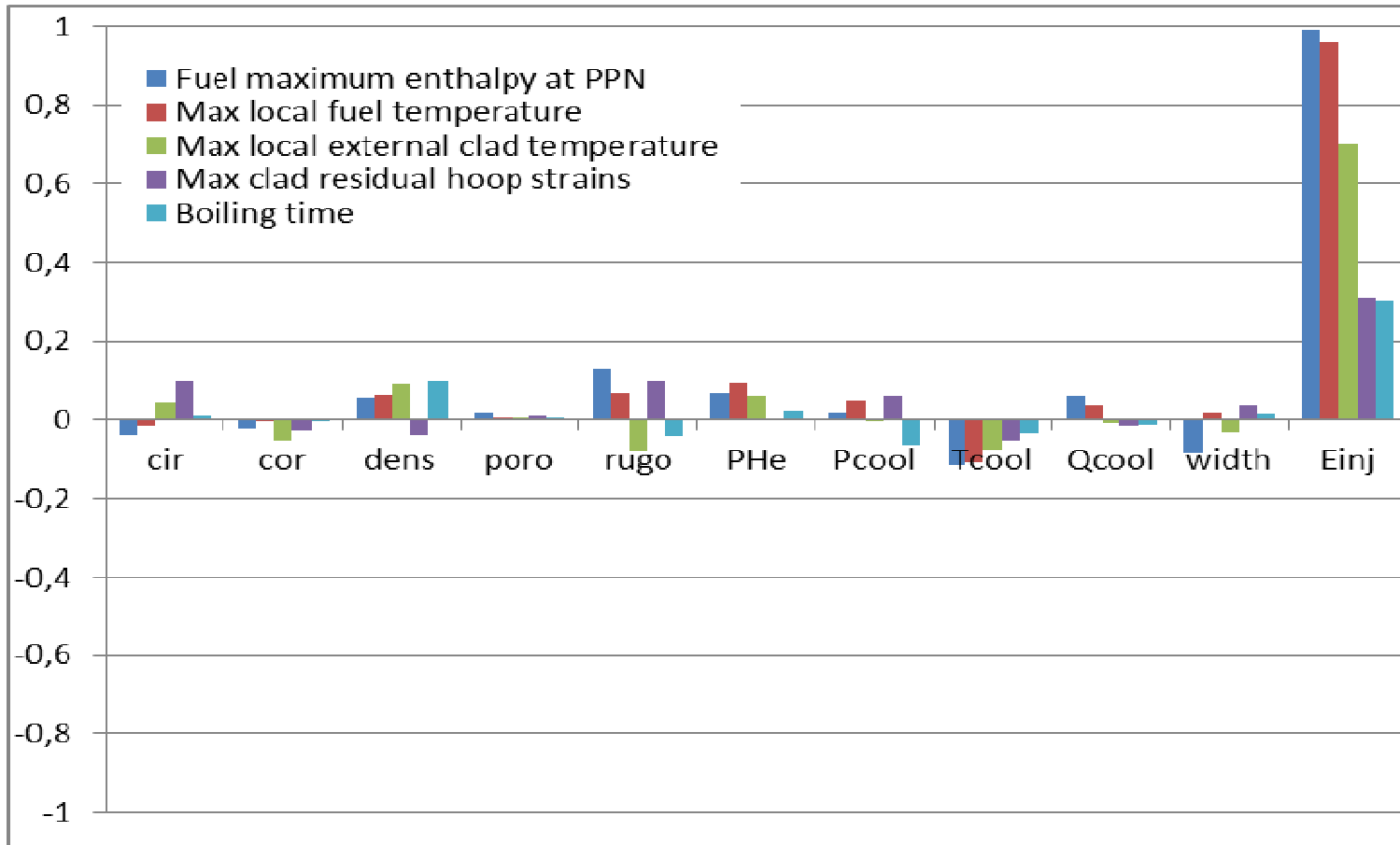
Exo7: même question pour $Y=X^2$ avec X d'espérance $=0$, de variance σ^2 et de loi symétrique.

$$\begin{aligned}\text{Cov}(X, Y) &= E((X - E(X))(X^2 - E(X^2))) \\ &= E(X^3) \\ &= 0\end{aligned}$$

Extension à un modèle non-linéaire monotone

- Régression sur les rangs: les valeurs des paramètres d'entrée et de la sortie sont remplacées par leurs rangs respectifs
- Mesures de sensibilité: coefficients de corrélation sur les rangs (même définition que ρ_j et ρ_j mais en remplaçant les valeurs par les rangs)
 - Coefficient de corrélation de Spearman sur les rangs (RCC)
 - Coefficient de corrélation partielle sur les rangs (PRCC)

Exemple: OECD RIA benchmark





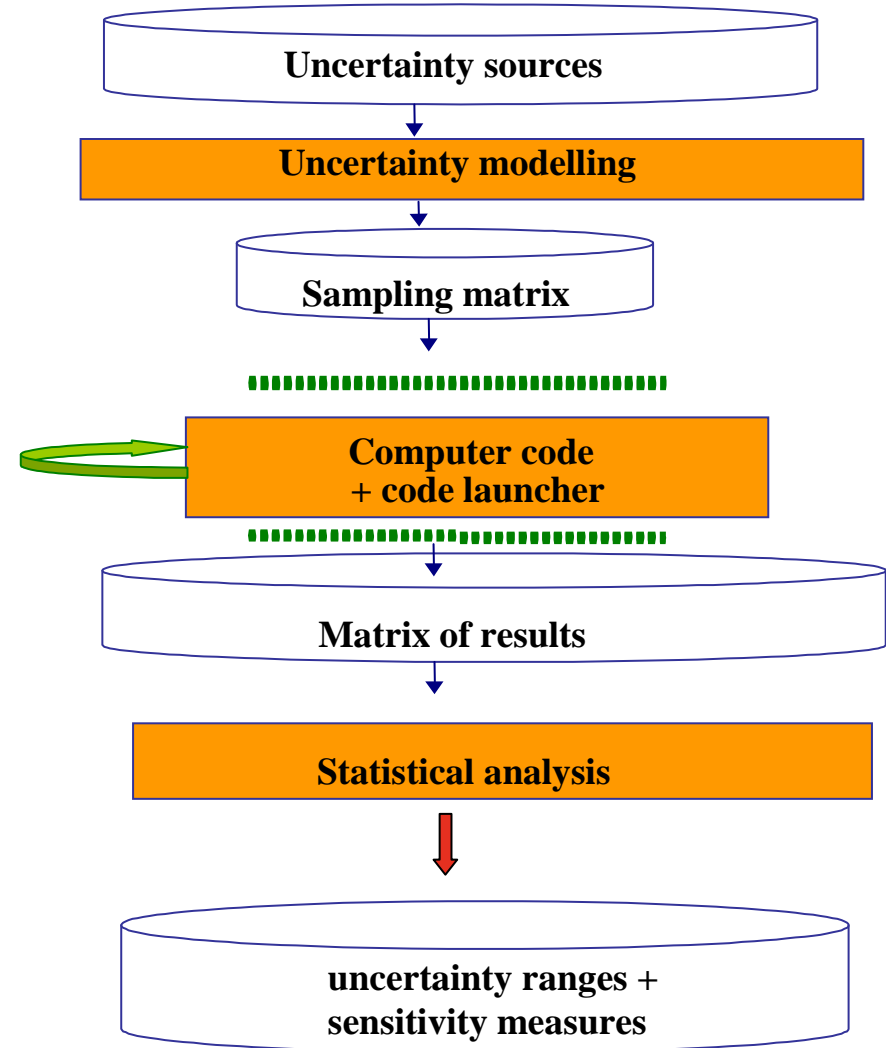
SUNSET : Statistical UNcertainty and Sensitivity Evaluation Tool



- Dispose d'une IHM
- Exécutable téléchargeable à partir de la forge SUNSET

<https://gforge.irsnn.fr/gf/project/sunset>

- OS: Windows, Unix, Linux
- Langage C++



Références

G. Saporta, *Probabilités, analyse des données et statistique*, Technip, 1990

JCGM 101:2008 Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method, 2008

J. Baccou, E. Chojnacki, Contribution of the mathematical modelling of knowledge to the evaluation of uncertainty margins of a LBLOCA transient (LOFT-L2-5), *Nuclear Engineering and Design*, 237 (19), 2064-2074, 2007.

E. Chojnacki, J. Baccou, S. Destercke, Numerical accuracy and efficiency in the treatment of epistemical and aleatory uncertainty, *International J. of General Systems*, 39(7), 683-704, 2010.

A. Saltelli, K. Chan, E.M. Scott, *Sensitivity Analysis*, Wiley Series in Probability and Statistics, 2000

B. Iooss, P. Lemaître, *A review on global sensitivity analysis methods*. In Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications, C. Meloni and G. Dellino (eds), Springer, 2015

G. Box and N. Draper, *Empirical model-building and response surfaces*. Wiley series in probability and mathematical statistics, 1987.

D. Benoist, Y. Tourbier, S. Germain-Tourbier, *Plans d'expériences : construction et analyse* Technique & Documentation - Lavoisier, 1994

S. Suard, S. Hostikka, J. Baccou, Sensitivity analysis of fire models using a fractional factorial design, *Fire Safety Journal*, 62:115-124, 2013