

Société de Calcul Mathématique SA

*Outils d'aide à la décision*

*depuis 1995*



La prise en compte des incertitudes dans les codes de calcul

Utilisation incorrecte de la Méthode de Wilks

par :

Bernard Beauzamy & Giovanni Bruna  
*PDG, SCM SA*                      *ancien Directeur Scientifique, IRSN*

mai 2024

## **I. Origine de nos travaux**

Ils trouvent leur source dans les contrats suivants :

- Framatome-ANP, 2003-2004 : Application de méthodes statistiques dans les analyses thermo-hydrauliques des études d'accident sur les réacteurs nucléaires (Luigi Catalani, Directeur Scientifique, Framatome) ;
- Institut de Radioprotection et de Sûreté Nucléaire : contrat-cadre 2003-2017 : méthodes probabilistes pour la sûreté nucléaire (Giovanni Bruna, Directeur Scientifique, IRSN) ;
- EDF SEPTEN, 2015 : Etudes relatives à la sûreté nucléaire.

## II. Les démonstrations de sûreté

Une démonstration de sûreté, pour un matériel, un équipement, un véhicule, une usine, consiste en l'analyse très minutieuse des risques avec, pour chacun, la réponse appropriée. Il s'agit de montrer que tous les événements fâcheux - ainsi que leurs initiateurs - ont été pris en considération et que l'on sait y faire face pour les éliminer, les maîtriser et en limiter l'impact environnemental dans le respect de la réglementation. Elle est normalement rédigée par l'Industriel qui propose le produit ; elle est ensuite validée par les Autorités de Sûreté de la branche correspondante. Typiquement, une démonstration de sûreté représente un gros volume de plusieurs centaines de pages et sa rédaction - en général assez complexe - prend du temps et comporte un coût économique et un investissement humain importants. Elle est revue et remise à jour périodiquement et, en tout cas, à chaque fois qu'un nouvel élément apparaît, et, bien entendu, à chaque fois qu'un accident se produit, ou aurait pu se produire.

Outre les événements, des modifications de référentiel interviennent : ainsi une installation jugée conforme (ou sûre) à une date peut ne plus l'être à une autre date ; si elle est jugée conforme dans un pays, elle peut ne pas être considérée telle dans un autre.

Il peut arriver qu'un accident se produise alors que toutes les précautions avaient été prises. Les Industriels cherchent à normaliser leurs process, c'est-à-dire à les rendre les plus constants et robustes possible. Mais la Nature impose des variabilités (sur les températures, les concentrations, les pressions, etc.) qui ne sont pas toujours prises en compte ; il peut en résulter des instabilités du process, qui n'avaient pas été anticipées.

La prise en compte des incertitudes dans les données pose un problème permanent. Dans beaucoup de cas, les seuls moyens d'investigation sont des codes de calcul disponibles sur le marché ou développés et validés pour un usage spécifique. L'exploitation de l'expérience de fonctionnement peut être difficile ou tout simplement impossible, et la réalisation d'expériences ciblées s'avère bien trop coûteuse ou carrément injustifiée du point de vue physique et/ou pratique. Se pose alors la question suivante : comment incorporer les incertitudes dans un code de calcul, qui, par définition, retourne un résultat dont la précision ne peut au mieux qu'être égale à celle des données d'entrée (en général elle est bien moindre) ? Peut-on affecter les résultats de marges jugées suffisantes pour tenir compte des incertitudes en entrée ?

## III. Code fin ou grossier ?

Ce n'est pas impossible, mais c'est d'autant plus difficile que le code est plus fin ; un code grossier sera mieux approprié. Prenons un exemple simple pour faire comprendre ceci.

On dispose d'un bureau en bois, de taille 2 m x 1 m, soit deux millions de mm<sup>2</sup>. On soupçonne que ce bureau peut être affecté par une fissure de taille 10 cm x 1 mm, soit 100 mm<sup>2</sup>. Si on utilise un outil d'investigation qui voit mm<sup>2</sup> par mm<sup>2</sup>, on aura énormément de mal à détecter la fissure et ce sera pratiquement impossible si on fait les essais au hasard. Par contre, si on utilise un outil d'investigation grossier, qui voit 10 cm x 10 cm, la détection est aisée et il suffit de 200 utilisations de l'outil. C'est du bon sens.

En réalité, les deux niveaux d'investigation sont nécessaires : l'approche grossière pour débroussailler le problème, l'approche plus fine pour une investigation plus approfondie une fois la première analyse réalisée.

## IV. Des runs lancés au hasard

En sûreté nucléaire, le code de calcul de référence est le code CATHARE (Code Avancé de ThermoHydraulique pour les Accidents de Réacteurs à Eau), code système de thermohydraulique diphasique développé depuis 1979 par le CEA dans le cadre d'un accord réunissant le CEA, EDF, FRAMATOME et l'IRSN (cf. <https://www.irsn.fr/recherche/code-cathare>). Il comporte environ 50 paramètres (cela dépend des versions) et chaque paramètre peut prendre au moins 10 valeurs (pour certains, la variation est continue) ; autrement dit, l'espace des configurations est de taille au moins  $10E50$ . Si on fait un milliard ( $10E9$ ) de runs, la proportion explorée n'est que  $10E(-41)$ , c'est infime. La probabilité d'atteindre la cible, c'est-à-dire de détecter une configuration dangereuse, est inférieure à celle qu'aurait un aveugle ivrogne de voir une fléchette lancée au hasard passer l'ouverture d'une porte entrebâillée.

En réalité, comme on le verra plus loin, si on veut faire les choses correctement, on ne déroulera jamais les études en faisant varier les paramètres à l'aveugle : on se laissera guider par la connaissance métier, qui permettra de se focaliser seulement et/ou en priorité sur les situations jugées potentiellement sensibles.

Les spécialistes du sujet, arcbutés comme il est d'usage sur leurs habitudes et leurs certitudes, diront : mais, pour nos runs aléatoires, nous utilisons une méthode probabiliste légitime, bien connue, la "méthode de Wilks". L'objet du présent article est de montrer de manière mathématique, c'est-à-dire au moyen d'un contre-exemple irréfutable, que l'utilisation de la méthode de Wilks peut s'avérer incorrecte. Pas légèrement incorrecte, mais fondamentalement incorrecte.

## V. Du Tribunal Correctionnel à la Cour d'Assises

On pourrait en rire, s'il ne s'agissait de sûreté nucléaire : dans beaucoup de domaines, les spécialistes sont catégoriques et les accidents se produisent tout de même ; il est arrivé qu'une digue s'effondre le jour même de la mise en eau.

Selon les cas, le Tribunal Correctionnel parlera de négligences, blessures ou homicides involontaires : le process était mal connu et toutes les précautions n'ont pas été prises (mais on ne sait pas exactement lesquelles il aurait fallu prendre).

Ici, les choses sont différentes : l'erreur est évidente et a été signalée. On peut admettre que, par ignorance, les responsables des études de sûreté continuent à mettre en œuvre des méthodes fautives : ils ne font que répéter des procédures qu'on leur a enseignées et qu'ils ont vues à l'œuvre un grand nombre de fois. Il en va différemment des Agences gouvernementales qui sont censées maîtriser ces sujets et avertir les Industriels. Si elles ne le font pas, leur responsabilité est engagée et, en ce cas, ce ne sera pas le Tribunal Correctionnel, mais la Cour d'Assises. Très vite, la justice risque d'être saturée !

## VI. Les incertitudes

Nous n'aborderons pas ici la question de la représentativité de ces codes de calcul : décrivent-ils correctement la réalité physique ? Cela ne peut résulter que d'une validation : il faut bien, un jour ou l'autre, faire une expérience physique, dont on comparera les résultats à ceux du code. Une telle validation est évidemment partielle : on n'est jamais complètement certain, quoi qu'on fasse, de la représentativité du code.

En particulier, il est pratiquement impossible de reproduire les conditions incidentelles en laboratoire. Les choses se compliquent encore quand on fait appel à la "sûreté passive" qui est censée opérer seulement à partir de phénomènes capables d'enclencher des contre-réactions qui contribuent au rétablissement des conditions de fonctionnement du système. De nombreux facteurs (de conception, d'échelle, d'environnement, de matériel, ...) empêchent tout simplement de reproduire, voire d'approcher, ces phénomènes.

La question que nous traitons ici est celle de la prise en compte des incertitudes : un code de calcul est évidemment déterministe ; on introduit des valeurs précises et il retourne un résultat précis. Si les données d'entrée sont soumises à incertitude, comme c'est toujours le cas en pratique, comment les prendre en compte ? La réponse des spécialistes est généralement : nous faisons de multiples "runs" du code (un "run", en terminologie anglo-saxonne, est une mise en œuvre du code), en faisant varier les données d'entrée, de manière à refléter les incertitudes.

### A. Approches scientifiques

Deux attitudes sont possibles :

#### 1. Approche déterministe régulière

Nous pouvons toujours normaliser les variables d'entrée pour qu'elles soient à valeurs entre 0 et 1 : il suffit de remplacer  $X$  par  $\frac{X - X_{\min}}{X_{\max} - X_{\min}}$ . Ceci fait, admettons que la valeur qui nous intéresse est  $X = 0.5$ , mais nous ne savons pas trop : ce pourrait être n'importe quoi entre 0.4 et 0.6.

Nous allons diviser l'intervalle  $[0.4-0.6]$  en (disons) 100 intervalles égaux et nous ferons un run pour chacun ; ainsi nous aurons une bonne perception de la réponse du code si la donnée n'est pas exactement 0.5.

Cette approche, appelée "approche déterministe régulière" est très satisfaisante, à ceci près que s'il se produit quelque chose de vraiment significatif entre 0.5111 et 0.5112, nous ne le verrons pas. Seul un expert du domaine, s'appuyant sur des considérations liées à la physique du process, saura que la zone aux alentours de 0.511 mérite une investigation particulière. C'est là le rôle de la connaissance métier, dont nous avons parlé plus haut. En d'autres termes, un code de calcul ne restitue que ce qu'on y a entré, sous une forme nouvelle, différente, certes plus accessible et exploitable. Il ne crée pas de la connaissance de façon spontanée. L'orientation doit venir de l'humain qui l'utilise. Si cette capacité d'orientation fait défaut, en particulier lorsque les

compétences se sont évaporées avec le temps, le recours à un traitement "à l'aveugle" n'est pas satisfaisant.

Un autre inconvénient de l'approche déterministe régulière est qu'elle est impraticable en grande dimension. Si le code dépend de 50 paramètres et que la zone d'intérêt couvre 10 valeurs pour chacun d'eux, l'espace à analyser comporte  $10^{50}$  valeurs, ce qui est hors de portée, quelle que soit la vitesse d'exécution du code.

Pour y remédier, on a souvent recours à une approche probabiliste :

## 2. Approche probabiliste

On va jeter au hasard un certain nombre de runs, en espérant qu'une certaine proportion d'entre eux permettra de mettre en évidence la zone sensible. Malheureusement, cette approche – pour répandue qu'elle soit – est fondamentalement erronée.

Commençons par des rappels simples. Si une variable aléatoire  $X$  suit une loi uniforme sur l'intervalle  $[0,1]$ , la probabilité qu'elle tombe dans un intervalle de largeur  $\alpha$ , inclus dans  $[0,1]$ , est proportionnelle à la taille de l'intervalle. Par conséquent, la probabilité qu'elle n'y tombe pas est  $1-\alpha$  et la probabilité qu'elle n'y tombe pas en  $n$  répétitions est  $(1-\alpha)^n$  ; ici les répétitions sont supposées indépendantes : le résultat qu'on obtient à chaque run du code ne dépend pas des précédents ni n'influe sur les résultats suivants.

Comme, pour  $\alpha > 0$  fixé,  $(1-\alpha)^n \rightarrow 0$  lorsque  $n \rightarrow +\infty$ , on peut à bon droit conclure que, si on fait un nombre de runs suffisant, viendra un moment où l'on tombera dans l'intervalle spécifié.

On peut même donner un énoncé quantitatif : soit  $\varepsilon > 0$ , on a  $(1-\alpha)^n < \varepsilon$  dès que

$n > \frac{\text{Log}(\varepsilon)}{\text{Log}(1-\alpha)}$  ; par conséquent, si par exemple  $\alpha = 1/100$  et  $\varepsilon = 0.1$ , cette inégalité sera satis-

faite dès que  $n \geq 230$ , autrement dit, si on prend un intervalle quelconque de taille  $1/100$  dans l'intervalle  $[0,1]$ , on a 99 chances sur 100, si l'on fait au moins 230 runs, de tomber au moins une fois dans cet intervalle.

Souvent, on introduit la notion de "niveau de confiance", et on donnera l'énoncé sous la forme : avec confiance d'au moins 99/100, la réitération de 230 runs du code de calcul permettra de détecter tout intervalle de taille 0.01.

Un tel énoncé, même s'il est simple sur le plan mathématique, est très difficile à comprendre, et encore plus difficile à accepter pour des tiers. Concrètement, pour bien comprendre, il faut se référer à la loi des grands nombres, et traduire l'énoncé sous la forme suivante : si nous répétons un million de fois l'expérience "faire 230 runs du code", alors 99/100 en moyenne donneront le résultat voulu, à savoir tomber au moins une fois dans l'intervalle considéré.

Sous cette forme, c'est encore pire ! La plupart des acteurs concernés, y compris les Autorités de Sûreté, ne sont pas familiers avec le niveau de confiance. De plus, ils n'ont aucun moyen de répéter l'expérience un million de fois.

C'est la raison pour laquelle il existe deux approches : l'approche déterministe qui ne veut pas entendre parler de probabilité mais de « certitude », non pas d'éviter les accidents (c'est irréaliste), mais d'en limiter ou ramener à zéro les conséquences sur l'environnement et les humains par la mitigation, et l'approche probabiliste. Les limites probabilistes dans les démonstrations de sûreté sont données par réacteur et par année d'exploitation. Avec des fonctionnements sur 60 ans et les parcs actuels, dans ces conditions une limite à  $10E(-6)$  équivaut à un événement majeur toutes les quelques années. Et encore faut-il savoir que ce  $10E(-6)$  signifie : ce n'est clair pour personne !

Nous sommes obligés de donner raison aux sceptiques : un énoncé à caractère probabiliste ne se comprend qu'au travers de la loi des grands nombres (répétitions virtuelles de l'expérience, en un nombre qui tend vers l'infini) ; elle n'a aucune valeur de certitude déterministe. Mais les probabilités font partie des lois de la Nature, qui elles-mêmes ne sont pas déterministes.

En 2015, à la demande de EDF/SEPTEN, la S.C.M. a réfléchi au moyen d'apporter des éléments de réponse à l'exigence théorique d'éliminer toutes les incertitudes au sein des démonstrations de sûreté. Sous cette forme abrupte, l'objectif est certainement impossible à attendre. Nous avons alors préconisé de repousser les incertitudes au-delà du périmètre du cahier des charges. Expliquons ceci sur un exemple concret : justifier la tenue d'une installation en cas de séisme.

On retiendra une magnitude de référence, par analyse de l'historique de ce qui s'est produit dans la zone, sur une longue période. Disons par exemple : aucun séisme de magnitude 4 ne s'est produit dans cette zone, depuis mille ans. Cela ne prouve en rien qu'un tel séisme est impossible dans l'avenir, mais l'argument "rien de tel ne s'est produit depuis mille ans" est de nature à rassurer l'opinion. Ensuite, par tous les moyens déterministes disponibles, nous montrons que "to the best of our knowledge", nos équipements vont résister à un séisme de magnitude 4 : ceci se fait par des calculs, des tests sur table vibrante, etc. Là encore, la démonstration n'est pas absolue, mais seulement "to the best of our knowledge". Comme on le constate, le cahier des charges est purement déterministe : réaliser des installations qui résistent à un séisme de magnitude 4, et l'incertitude sur l'hypothèse (qui porte sur la magnitude 4 du séisme) a été rejetée hors du cahier des charges.

### **3. La "méthode de Wilks"**

Elle consiste en un énoncé probabiliste extrêmement simple (niveau Terminale), traduisant les formules vues plus haut. Elle s'énonce ainsi : étant donné un phénomène aléatoire quelconque, si on se fixe des boîtes d'égale probabilité, on finira certainement par tomber dans chacune d'entre elles, au bout d'un nombre suffisant de "runs" ou d'expérimentations. Cette méthode est souvent utilisée dans les études probabilistes de sûreté, parce qu'elle permet de déterminer le nombre minimum de runs nécessaires (pour une expérience, pour un code de calcul), afin d'évaluer un risque avec un niveau de sécurité jugé suffisant. L'énoncé précis est ([Wilks]) :

La méthode de Wilks est la méthode la plus employée à ce jour en thermo-hydraulique nucléaire pour estimer un quantile. Elle permet de déterminer le nombre minimum  $N$  de simulations nécessaires à l'obtention d'un majorant de la valeur d'un quantile  $T_{\alpha,\beta}$ . Dans cette notation,  $\alpha$  désigne le niveau de confiance attendu ("nous sommes certains à 95%") et  $\beta$  désigne la proportion de résultats prise en considération ("les 5% de températures les plus élevées").

Ce nombre est donné par la formule de Wilks, qui résulte directement de l'expression de l'intervalle de confiance d'une loi binomiale :

$$1 - \sum_{i=N-r+1}^N \binom{N}{i} p^i (1-p)^{N-i} \geq \beta$$

$p$  est la probabilité qu'une simulation quelconque aboutisse à un résultat satisfaisant (inférieur à la valeur critique). L'estimation du quantile  $T_{95,95}$  est donnée par la valeur pire cas obtenue au cours de  $N$  simulations avec  $N$  donné par les conditions :

$$1 - \alpha^N \geq \beta, \text{ soit } \alpha^N \leq 1 - \beta, \text{ ou } N \geq \frac{\text{Ln}(1-\beta)}{\text{Ln}(\alpha)}.$$

Numériquement, avec  $\alpha = 0.95$  (confiance à 95%) et  $\beta = 0.95$ ,  $1 - \beta = 0.05$  : 5% des températures les plus élevées, on obtient :

$$N \geq \text{Ln}(0,05) / \text{Ln}(0,95) \approx 58.4. \text{ Soit } N \geq 59.$$

Démonstration de la formule

Soit  $X$  une variable aléatoire à valeurs réelles. Divisons l'ensemble des valeurs prises par  $X$  (peu importe en quoi elles consistent) en  $n$  sous-ensembles de même probabilité (en anglais "bins", souvent traduit par "boîtes/bacs"). Evidemment, la probabilité de chaque boîte est  $p = \frac{1}{n}$ .

Désignons par  $B_1, \dots, B_n$  ces boîtes. Choisissons l'une d'entre elles, par exemple la dernière,  $B_n$ . Supposons que nous fassions  $N$  tirages indépendants de la variable  $X$  (peu importe quelle est la loi de cette variable). Alors la probabilité de ne jamais tomber dans  $B_n$  au cours de ces  $N$  tirages est :

$$q = (1-p)^N.$$

Nous avons déjà vu cela plus haut : la probabilité de ne pas tomber dans  $B_n$  lors d'un tirage est  $1-p$  et donc la probabilité de ne jamais y tomber en  $N$  tirages est  $(1-p)^N$ , puisque les tirages sont indépendants.

Comme  $(1-p)^N \rightarrow 0$  lorsque  $N \rightarrow \infty$ , nous sommes de plus en plus sûrs de toucher toutes les boîtes, lorsque le nombre de tests augmente. Fixons-nous un seuil  $\beta$ , traditionnellement 95% ; nous constatons que notre probabilité de toucher la dernière boîte en  $N$  runs dépasse ce seuil dès que :

$$(1-p)^N \leq 1-\beta.$$

On retrouve la notation de la formule en posant  $\alpha = 1-p$ . Avec les valeurs numériques prises ici, à savoir  $p = \frac{1}{20}$ , on a  $\alpha = 0.95$ , et avec  $\beta = 0.95$ , on trouve effectivement  $N \geq 59$ .

#### 4. Application de la méthode à l'analyse de sûreté

Celle-ci est généralement faite sous la forme suivante.

On dispose d'un code de calcul, supposé représenter un phénomène complexe. Nous prendrons l'exemple du code de thermohydraulique "CATHARE" (Code Avancé de THERmohydraulique pour les Accidents des Réacteurs à Eau, développé par Areva, le CEA, EDF et l'IRSN), mentionné précédemment. Ce code est utilisé dans les calculs de transitoire, entre autres pour décrire le comportement d'un réacteur REP (Réacteur à Eau Pressurisée) en cas de grosse brèche dans le circuit primaire ; il calcule la variation de la température du réfrigérant présent dans ce circuit en fonction du temps. Le code comporte un nombre très élevé de paramètres (40 ou davantage, selon les versions). Il est extrêmement complexe, et a requis des dizaines d'années de développement. Le code est accepté par les Autorités de Sûreté françaises dans les démonstrations de sûreté, moyennant les éléments de validation d'usage.

Dans le livre Zeydina-Beauzamy "Probabilistic Information Transfer" [PIT], nous prenons l'exemple de ce code pour montrer comment "propager" l'information, de manière probabiliste, à partir de situations où les runs ont été faits à des situations où aucune information n'est connue.

La question fondamentale, pour les analyses de sûreté, est donc celle-ci : on dispose d'un petit nombre de runs (du code, de l'expérience), mettons 1 000 pour fixer les idées, et on voudrait en déduire une réponse à une question liée à la sûreté. Par exemple, pour CATHARE, un seuil est fixé arbitrairement à 1 200°C ; admettons que nos 1 000 runs aient tous donné des températures inférieures, voire bien inférieures, cela prouve quoi ? Que pouvons-nous remettre aux Autorités de Sûreté ?

La méthode de Wilks est utilisée, pour répondre à cette question, de la manière suivante :

Nous avons nos 40 paramètres. Sur chacun d'entre eux, nous choisissons arbitrairement une loi de probabilité. Par exemple, sur le premier, nous choisirons une loi uniforme sur tel intervalle ; pour le second nous choisirons une loi de Gauss de moyenne tant et de variance tant, et ainsi de suite jusqu'au 40<sup>ème</sup>.

Par exemple, pour le paramètre  $X_{16}$  "pression accu", on retient une loi uniforme entre  $m = 4\,137$  kbars et  $M = 4\,385$  kbars. Ceci n'a rien de choquant : cela signifie que, les experts ne connaissant pas la valeur exacte que peut prendre ce paramètre, lui attribuent n'importe quelle valeur entre ces bornes, avec égale probabilité. Cela traduit le fait que la connaissance physique du phénomène n'est pas parfaite.

Comme il a été dit précédemment, ces lois sont établies "à dire d'expert", c'est-à-dire que les experts considèrent que, en cas d'accident, le premier paramètre sera effectivement compris entre  $a$  et  $b$ , que le second sera plutôt concentré autour d'une certaine valeur, etc. Ces affirmations sont assurément discutables, mais elles ne sont pas absurdes et elles peuvent servir de base au raisonnement. Notre critique ne porte pas sur ce point.

Après quoi, on procède à des tirages aléatoires de chaque paramètre selon sa loi propre ; si pour le premier, on a choisi une loi uniforme, on fait un tirage selon une loi uniforme. Si pour le second on a choisi une loi de Gauss, on fait un tirage selon une loi de Gauss, et ainsi de suite pour les 40 paramètres. On répète l'ensemble de la procédure par exemple 1 000 fois (cela fait donc au total 40 000 tirages). On dispose ainsi de 1 000 "runs", chacun correspondant à une situation de fonctionnement du code (une situation de fonctionnement requiert 40 paramètres). Pour chaque run, le code fonctionne et retourne une température.

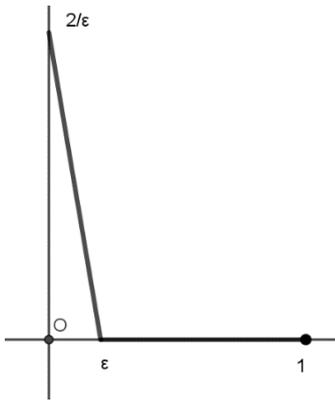
Les experts prétendent alors que, puisqu'on a fait plus de 59 runs choisis aléatoirement, la méthode de Wilks nous dit que nous avons plus de 95 chances sur 100 d'avoir détecté le quantile 95 des températures les plus élevées. Autrement dit, que nous sommes quasiment certains de la représentativité de nos essais : nos runs vont révéler la température la plus élevée.

Cette affirmation est entièrement fautive. Nous allons le voir sur un exemple simple.

Supposons que le code CATHARE, noté  $CT$ , ne dépende que du seul paramètre  $x = X_{16}$  (pression accu), que nous normaliserons pour varier entre 0 et 1 : c'est  $x = \frac{X_{16} - 4137}{4385 - 4137}$ . Rappelons que le code retourne une température. Nous le notons  $CT(x)$ . Par hypothèse, la forme exacte de cette fonction n'est pas connue (elle l'est de ceux qui ont conçu le code, et la forme de la fonction dépend certainement de tous les autres paramètres).

Supposons que la forme exacte du code soit une fonction triangle :

$$y = \frac{-2}{\varepsilon^2}x + \frac{2}{\varepsilon} \text{ si } 0 \leq x \leq \varepsilon, \quad y = 0 \text{ si } \varepsilon \leq x \leq 1.$$



Le graphe est donné dans la figure ci-contre. Les températures, en ordonnées, vont de 0 à  $\frac{2}{\varepsilon}$ . Le quantile à 95% le plus élevé va de  $\frac{95}{100} \frac{2}{\varepsilon}$  à  $\frac{2}{\varepsilon}$ , pour les températures, et il est obtenu pour  $0 \leq x \leq \frac{5}{100} \varepsilon$ .

Prenons, pour fixer les idées,  $\varepsilon = \frac{1}{1000}$ . Les températures iront donc jusqu'à 2 000°C, largement supérieur à la limite acceptable, qui est 1 200°C. L'intervalle en  $x$  où ceci se produit est  $0 \leq x \leq 5 \times 10^{-5}$ . Cet intervalle est microscopique, et nous n'avons aucune chance de le pénétrer si nous faisons 59 runs aléatoires entre 0 et 1.

Avec  $l = 5 \times 10^{-5}$ , la probabilité de ne pas tomber dans l'intervalle en un run est  $1-l$ , la probabilité de ne pas y tomber en  $N$  runs est  $(1-l)^N$  et la probabilité d'y tomber en  $N$  runs est  $p = 1 - (1-l)^N$ .

Avec  $N = 59$ , on trouve  $p \approx 0.0029$  et avec  $N = 1000$ , on trouve  $p \approx 0.049$  : on est loin du compte !

D'où vient l'erreur ? L'estimation de Wilks serait correcte, si elle était appliquée à l'intervalle  $0 \leq x \leq \varepsilon$ . Elle ne l'est pas, si elle est appliquée à tout l'intervalle  $0 \leq x \leq 1$ . Or, dans la pratique, nous n'avons aucune information qui permette de restreindre l'intervalle d'étude à  $0 \leq x \leq \varepsilon$ , pour la bonne raison que  $\varepsilon$  n'est pas connu. Nous sommes donc obligés, par principe, de tester tout l'intervalle  $0 \leq x \leq \varepsilon$ , mais si les valeurs élevées de la température se réfugient dans un "coin" de l'intervalle d'étude, nous ne les trouverons jamais par un nombre fixé de runs, car la taille de l'intervalle critique n'est pas connue d'avance.

Reprenons l'énoncé de Wilks, vu plus haut, pour bien comprendre ce qui cloche : *soit  $X$  une variable aléatoire à valeurs réelles. Divisons l'ensemble des valeurs prises par  $X$  (peu importe en quoi elles consistent) en  $n$  sous-ensembles de même probabilité. Evidemment, la probabilité de chaque boîte est  $p = \frac{1}{n}$ .*

Mais ceci est impossible dans l'exemple vu plus haut : nous avons une boîte (à savoir  $CT = 0$ ) dont la probabilité est très voisine de 1, à savoir  $1 - \varepsilon$  et il ne reste que  $\varepsilon$  à partager entre toutes les valeurs non-nulles de la température.

### 5. Cas de la dimension élevée

Les choses ne font qu'empirer lorsque la dimension s'élève. Si on regarde d'abord le cas simple où le code ne dépend que de deux paramètres, mettons  $CT = X_1 + X_2$ , où  $X_1$  et  $X_2$  suivent toutes deux une loi uniforme et sont indépendantes, la réponse du code n'est pas une loi

uniforme mais une loi triangle. La loi réponse se concentre de plus en plus lorsque la dimension augmente.

Voici maintenant un exemple explicite où les valeurs élevées de la température ne pourront jamais être détectées par la méthode de Wilks, telle qu'elle est utilisée actuellement. Ce contre-exemple, complètement évident, met bien en évidence l'erreur commise.

On suppose que la température, dépendant de 40 paramètres, est donnée par la formule :

$$CT = 2000 x_1 x_2 \cdots x_{40}$$

où chacun des  $x_i$  suit une loi uniforme sur l'intervalle  $[0,1]$ . La valeur maximale de la température est  $2000^\circ\text{C}$ , et elle est obtenue si tous les  $x_i$  valent 1. On constate que le code est linéaire par rapport à chaque paramètre pris séparément et qu'il est très régulier (aucune discontinuité d'aucune sorte).

Il est pratiquement impossible, par quelque simulation numérique que ce soit, de mettre en évidence une température supérieure à  $1\,000^\circ\text{C}$ . En effet,  $CT > 1\,000$  équivaut à :

$$x_1 x_2 \cdots x_{40} > \frac{1}{2} \tag{1}$$

La propriété (1) implique que nécessairement, dans ce cas, tous les  $x_i$  doivent être  $> \frac{1}{2}$ . Or :

$$P\left(x_i > \frac{1}{2}, \forall i\right) = \frac{1}{2^{40}} \approx 0.9 \times 10^{-12} \tag{2}$$

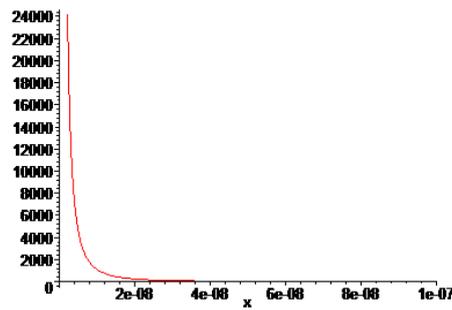
Il faudra donc de l'ordre de 1000 milliards de runs pour rencontrer une situation où tous les  $x_i$  seront  $> \frac{1}{2}$ .

Ceci se comprend très bien intuitivement : si l'on tire 40 fois au hasard selon une loi uniforme entre 0 et 1, il est très vraisemblable que l'un des tirages au moins sera entre 0 et 1/2, et en ce cas le produit le sera aussi. La détection de la situation à risque est pratiquement impossible par une simulation numérique.

**Remarque.** – On peut calculer précisément la probabilité de l'événement (1). En effet, le produit de  $n$  variables aléatoires indépendantes, suivant une loi uniforme sur  $[0,1]$ , a pour densité (voir [Dettmann – Georgiou]) :

$$f(x) = \frac{1}{(n-1)!} \left( \text{Log} \left( \frac{1}{x} \right) \right)^{n-1}$$

Voici l'aspect du graphe de la densité de probabilité, pour  $n = 40$  :



## 6. En conclusion

Un code de calcul comportant la prise en compte de 40 paramètres, qui a nécessité des dizaines d'années de développement, est un objet fondamentalement complexe. Il est naïf et illusoire de croire que, en lançant 1 000 runs, voire un million de runs au hasard, Dieu aura la bonne volonté de guider nos fléchettes vers les endroits que nous souhaitons atteindre.

Si le code comporte 40 paramètres et que chacun est discrétisé en 10 valeurs possibles, nous obtenons  $10^{40}$  possibilités. L'exploration par 59 runs, ou par 1000, ou par un million, reste de toute façon infime, qu'on le veuille ou non, qu'on utilise Wilks ou non. Nous avons, avec nos fléchettes, des milliards de fois moins de chances d'atteindre notre cible que n'en aurait un aveugle ivre lançant une fléchette au hasard ; il espère qu'elle parviendra à franchir une porte à peine entrebâillée.

## 7. Recommandations

Il est fondamentalement absurde et malsain de s'en remettre au hasard pour l'exploration d'un phénomène que nous ne comprenons pas. Bien au contraire, il est nécessaire de combiner l'expertise physique et l'exploration probabiliste locale, pour la valider. Concrètement, il faudra :

- Définir des scénarios, par exemple selon les conditions d'exploitation du réacteur ;
- Pour chaque scénario, définir des conditions accidentelles ;
- Pour chacune, par expertise, mettre en évidence les paramètres les plus influents ;
- Eliminer les paramètres les moins importants et se concentrer sur les paramètres les plus significatifs (voir notre livre [NMP]) ;
- Valider ce choix en procédant à des tirages aléatoires au voisinage des situations retenues ;
- Eliminer les paramètres qui sont sans influence, pour réduire la combinatoire du problème ; après avoir ainsi réduit la dimension, rechercher les configurations à risque, et concentrer les runs sur celles-là, pour les identifier le mieux possible ;
- Répéter ces raisonnements de manière à avoir la meilleure connaissance possible des paramètres significatifs et des situations critiques : concentrer les runs à la seconde génération sur les situations reconnues comme critiques à la première, et ainsi de suite.

On peut résumer ceci de manière simple : avant de passer un calcul, effectuez suffisamment de dégrossissages, d'analyses paramétriques et de sensibilité, et surtout faites travailler votre matière grise, afin d'avoir un ordre de grandeur dans votre tête avant de lire les résultats des calculs.

Malheureusement, plus personne ne fait cela : la facilité est telle que l'on lance les calculs sans réfléchir... et on se fie à la réponse de l'ordinateur comme s'il s'agissait d'un oracle.

Pour plus d'information, consulter les références [Bruna1], [Bruna2], [Bruna3], [Bruna4], [Bruna5].

## 8. Références

[Bruna1] G.B. Bruna et alii, “Evidence-based background for constrained uncertainty quantification in a core transient analysis”, *Annals of Nuclear Energy*, Volume 164, 15 December 2021, 108606.

[Bruna2] G.B. Bruna, “Prendre la mesure du risque : du risque mesuré aux limites de la mesure”, in S. Bretesché, C. Harpet, S. Ollitrault V. Héquet, (Coord.), « Le risque environnemental. Entre sciences physiques et sciences humaines ». Presses des Mines. Transvalor, Paris, 2019,

[Bruna3] G.B. Bruna, “Uncertainty in design and operation. How dealing with?” Plenary lecture provided at the Best Estimate Plus Uncertainty (BEPU) International Conference 2018 - Multi-Physics Multi-Scale Simulation with Uncertainty, Lucca, May 13 -19, 2018.

[Bruna4] G.B. Bruna et al., “Best Estimate plus Uncertainty (BEPU): why it is still not widely used?”, *Nucl. Techn.*, vol. 205, 2019 - Issue 12: Selected papers from the 2018 Best Estimate Plus Uncertainty International Conference (BEPU 2018), pp. 1529-1539, published online: 15 Jul 2019.

[Bruna5] G.B. Bruna, “Du risque et de sa perception”, *Variances*, 2020, <http://variances.eu/?p=5246>.

[Dettmann – Georgiou] Carl P. Dettmann and Orestis Georgiou : Product of n independent Uniform Random Variables, *Statistics and Probability Letters*, 79 (2009) 2501–2503.

[NMP] Bernard Beauzamy : *Nouvelles Méthodes Probabilistes pour l'évaluation des risques*. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN 978-2-9521458-4-8. ISSN 1767-1175, avril 2010.

[PIT] Olga Zeydina et Bernard Beauzamy : *Probabilistic Information Transfer*. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN: 978-2-9521458-6-2, ISSN : 1767-1175, mai 2013.

[WILKS] WILKS, S. S., Determination of Sample Sizes for Setting Tolerance Limits, *The Annals of Mathematical Statistics*, Vol.12, pp. 91-96, 1941.