

ESTIMATION D'UN QUANTILE CONCOURANT A LA MAITRISE D'UN DIMENSIONNEMENT THE BEST QUANTILE ESTIMATION TO VALIDATE DIMENSIONING

André Cabarbaye
CNES / CAB INNOVATION
3 rue de la Coquille – 31500 Toulouse
Tél. 05 61 54 68 08
andre.cabarbaye@cnes.fr /
andre.cabarbaye@cabinnovation.fr

Roland Laulheret
Centre National d'Etudes Spatiales (CNES)
18 av. Edouard Belin – 31401
Tél. 05 61 27 47 19
roland.laulheret@cnes.fr

Résumé

Cherchant à valider des dimensionnements de nature probabiliste à partir de résultats de simulation stochastique, cet article teste deux méthodes d'estimation des quantiles : la méthode de Wilks, qui est traditionnellement utilisée en thermo-hydraulique nucléaire, et la méthode du Bootstrap, qui est préconisée par un groupe méthodologique de l'OCDE pour démontrer qu'en cas de perte de réfrigérant du circuit primaire d'un réacteur nucléaire, la température de gaine des crayons combustibles ne dépassera pas une température critique dans 95 % des cas.

Pour tester les deux méthodes, nous avons considéré un phénomène physique régi par une loi théorique donnée dont on connaît a priori la valeur des quantiles. Nous avons alors estimé ces mêmes quantiles par les deux méthodes à partir d'un grand nombre d'échantillons générés par simulation de Monte-Carlo de la loi théorique. La méthode de Wilks donne des estimations en accord avec la confiance attendue contrairement à la méthode du Bootstrap qui doit être rejetée pour estimer un quantile même si son intervalle de confiance semble resserré. En effet, l'estimateur du quantile obtenu par cette méthode ne correspond pas à celui de la population mère mais à celui d'une population virtuelle générée à partir d'un unique échantillon.

Summary

Seeking to validate designs from the results of stochastic simulation, this article tests two quantile estimation methods: Wilks method, which is traditionally used in nuclear thermo-hydraulic, and the Bootstrap method, which is advocated by a group of the OCDE to demonstrate that in case of loss of primary coolant circuit in a nuclear reactor, the temperature will not exceed a critical temperature in 95 % of cases.

To test the two methods, we considered a physical phenomenon governed by a probability law which we know the quantile value. Then we tried to estimate this quantile value by both methods from a large number of samples generated by Monte-Carlo simulation from the probability law. The method of Wilks provides estimates consistent with the expected confidence as opposed to the Bootstrap, which must be rejected although his confidence interval appears tight. Indeed, the estimator of quantile obtained by this method does not match that of the parent population, but to that of a virtual population generated from one sample.

Introduction

Les codes de calcul interviennent de plus en plus dans le dimensionnement des systèmes. Aussi, la maîtrise de leurs incertitudes constitue une préoccupation majeure des concepteurs ainsi que des organismes de sûreté et de certification. Ces incertitudes ont des origines multiples : méconnaissance des phénomènes physiques, limitation et représentativité des données expérimentales, dispersion et corrélation éventuelles des données d'entrée, faiblesse des modèles numériques, etc.

Pour pallier cette difficulté, un conservatisme plus ou moins maîtrisé est introduit dans les calculs selon le degré de connaissance et l'ampleur des risques encourus. Dans le domaine spatial, par exemple, les satellites sont le plus souvent dimensionnés pour assurer une mission de référence dans les pires cas de fonctionnement (en fin de vie, durant les éclipses, etc.). Cette approche simplifie la validation des systèmes en limitant les vérifications à un cas de référence unique au moyen de simulateurs déterministes. Mais elle a aussi pour conséquences un surdimensionnement général des produits par rapport à leur utilisation réelle, la non utilisation de ressources disponibles et l'impossibilité de bénéficier de certains effets de seuil. Pour un satellite, par exemple, ce surdimensionnement se caractérise par une capacité énergétique surabondante en début de vie (dimensionnée pour compenser les dégradations et usures subies pendant toute la mission) qui n'est pas utilisée pour accroître les capacités opérationnelles du système (nombre d'images prises par un satellite d'observation de la Terre par exemple) et l'impossibilité éventuelle d'utiliser un lanceur de plus faible capacité beaucoup moins onéreux avec lequel l'essentiel de la mission pourrait pourtant être couvert.

Une approche dite « Best Estimate » cherche, par ailleurs, à décrire au mieux la réalité en fonction de l'état de l'art, afin de mieux évaluer les risques tout en évitant les surdimensionnements superflus. Toujours dans le domaine spatial, une mission probabiliste de référence pourrait être négociée avec le client comme expression du besoin (nombre moyen d'activations, zones concernées, durées...). Cette mission serait exécutée tant que les ressources à bord restent disponibles (niveau de décharge batterie, prévision d'éclipses...) mais pourrait ne pas être assurée en pires cas, dans la mesure où l'objectif de disponibilité requis est globalement atteint pendant toute la durée de la mission. La validation du système serait réalisée de manière probabiliste au moyen de simulateurs stochastiques permettant de considérer les aléas de toute nature portant sur la mission, les conditions orbitales et l'état du système à l'instant courant (défaillances et dégradations à bord).

Mais comment définir le nombre de simulations nécessaires à la démonstration de la tenue des objectifs ? C'est pour répondre à cette question que nous nous sommes intéressés à la méthode de Wilks, qui est traditionnellement utilisée en thermo-hydraulique nucléaire, et à celle du Bootstrap, qui est actuellement préconisée par un groupe méthodologique de l'OCDE (Programme BEMUSE [3]) pour démontrer qu'en cas de perte de réfrigérant du circuit primaire d'un réacteur nucléaire, la température de gaine des crayons combustibles ne dépassera pas une température critique dans 95 % des cas.

Le quantile α à 95% peut être estimé par ces deux méthodes avec une confiance β elle-même égale à 95% et doit être inférieur à la température critique ($T_{95,95} < T_c$), comme l'illustre la figure 1 (on note $T_{\alpha,\beta}$ le quantile α au taux de confiance β).

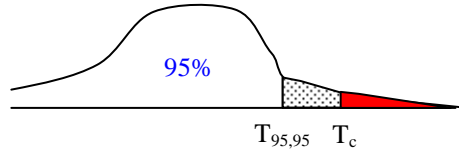


Figure 1. Démonstration de la tenue d'un objectif par estimation d'un quantile

A partir d'un échantillon original, correspondant dans notre cas aux résultats d'un certain nombre de simulations réalisées, la méthode du Bootstrap consiste à générer un grand nombre d'échantillons fictifs, par tirage aléatoire avec remise. L'objectif est ainsi de mieux exploiter l'information contenue dans l'échantillon original même si aucune information nouvelle n'est ainsi créée.

Mais si ce sur échantillonnage permet d'améliorer certains estimateurs par rapport à ceux obtenus directement à partir de l'échantillon initial, nous doutions toutefois qu'il puisse améliorer significativement la précision d'un quantile à 95% et surtout autoriser une réduction de la taille de l'échantillon initial pour un même objectif à atteindre. Aussi avons-nous cherché à tester la capacité réelle de ces deux méthodes à l'estimation d'un quantile.

Le but de cette communication est de présenter les résultats obtenus au terme de cette évaluation afin de pouvoir confronter nos conclusions sur ces aspects méthodologiques, inhérents à tout dimensionnement probabiliste, avec celles des spécialistes d'autres domaines dont notamment ceux du secteur nucléaire.

1. Méthodes

1.1. Méthode de Wilks

La méthode de Wilks [1] 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 $T_{\alpha,\beta}$. Ce nombre est donné par la formule de Wilks, qui résulte directement de l'expression de l'intervalle de confiance d'une loi binomiale :

$$1 - \sum_{N-r+1}^N C_n^i \alpha^i (1-\alpha)^{N-i} \geq \beta$$

α est la probabilité qu'une simulation quelconque aboutisse à un résultat satisfaisant (inférieur à la valeur critique), et r est le rang des pires cas obtenus durant toutes les simulations. L'estimation du quantile $T_{95,95}$ est donnée par la valeur pire cas obtenue au cours de N simulations ($r = 1$) avec N donné par l'expression :

$$1 - \alpha^N \geq \beta \text{ soit } \alpha^N \leq \beta - 1 \text{ ou } N \geq \ln(\beta - 1) / \ln(\alpha)$$

$$N \geq \ln(0,05) / \ln(0,95) = 58,4039748 \quad N \geq 59$$

Ce même quartile $T_{95,95}$ peut être estimé par la pire des valeurs obtenue à l'exclusion du pire cas ($r = 2$) avec N donné par l'expression :

$$1 - (N \alpha^N (1 - \alpha) + \alpha^N) \geq \beta \text{ soit } N \geq 93$$

ou par la pire des valeurs obtenue à l'exclusion des 2 valeurs pires cas ($r = 3$) soit $N \geq 124$

1.2. Méthode du Bootstrap

Proposée à la fin des années 70 par Bradley Efron [2], le Bootstrap est une méthode de sur échantillonnage qui consiste à créer un grand nombre d'échantillons, par tirage aléatoire avec remise, à partir d'un échantillon original, sans recourir à de nouvelles observations (figure 2).

```

2 5 7 8 9
  ↓
2 2 7 9 9
5 8 8 9 9
2 5 7 8 8
. . . . .
    
```

Figure 2. Exemple d'échantillons générés par Bootstrap

Même si aucune information nouvelle n'est ainsi créée, on cherche à exploiter toute l'information contenue dans l'échantillon original. A titre d'illustration, les distributions de la figure 3 représentent deux échantillons de 59 simulations dont la plus grande valeur (pire cas) est utilisée pour estimer le quantile $T_{\alpha,\beta}$ par la méthode de Wilks. A la vue de la forme des distributions, on peut supposer que l'estimation issue de l'échantillon 2 est plus conservatrice que celle issue de l'échantillon 1.

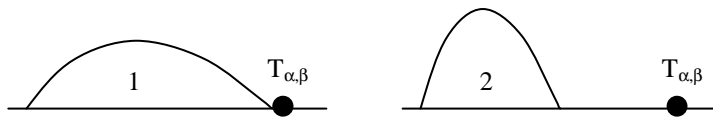


Figure 3. Distribution d'échantillons utilisée pour estimer un quantile

A partir d'un échantillon original, une famille d'échantillons est créée par Bootstrap sur laquelle on effectue des estimations de paramètres divers comme s'il s'agissait d'autant d'échantillons tirés de la population mère. Ainsi dans l'exemple de la figure 4, la moyenne de la population peut être estimée par la moyenne des valeurs moyennes obtenues sur chacun des échantillons générés par Bootstrap et l'écart-type calculé sur ces mêmes valeurs donne une estimation de ce qui est appelé l'erreur standard du paramètre étudié.

Echantillon original :	2	5	7	8	9	Moyenne	6,2
	↓						
Echantillons générés par Bootstrap	2	2	7	9	9		5,8
	5	8	8	9	9		7,8
	2	5	7	8	8		6
						Moyenne :	6,53
						Erreur standard (écart-type) :	1,10

Figure 4. Estimation d'une moyenne par Bootstrap

On procède de la même manière pour estimer un quantile et l'erreur standard correspondante, puis un intervalle de confiance, par application du théorème central-limite, en considérant que la distribution des quantiles est approximativement normale.

1.3. Méthode de test

La démarche de validation choisie pour tester les deux méthodes d'estimation d'un quantile est simple et pragmatique. Nous avons considéré un phénomène physique régi par une loi théorique donnée dont on connaît a priori la valeur des quantiles. Nous avons alors estimé ces mêmes quantiles par les deux méthodes à partir d'un grand nombre d'échantillons générés par simulation de Monte-Carlo de la loi théorique. L'outil SIMCAB de la société CAB INNOVATION a pour cela été utilisé et les fichiers de calcul sont accessibles par Internet [4].

Ce test a été mené avec deux lois théoriques : une loi normale de moyenne 800 et d'écart-type 100 et une loi lognormale de paramètres $\mu = 5$ et $\sigma = 1$. La valeur des quantiles correspondant à 95 % se calcule directement sous Excel, soit respectivement 964 et 769.

2. Résultats

2.1. Méthode de Wilks

Pour valider cette méthode, on tire aléatoirement un échantillon de 59 valeurs distribuées suivant la loi théorique et on vérifie que la plus grande valeur obtenue parmi ces 59 valeurs est bien inférieure à la valeur théorique du quantile à 95% dans une proportion d'environ 95 %.

La fonction de répartition de l'écart entre la valeur maximale de l'échantillon et la valeur théorique du quantile, présentée aux figures 5 et 6, a été obtenue après 2000 simulations d'un échantillon de 59 valeurs distribuées respectivement selon les lois normale et lognormale. Dans 95 % des cas environ, l'estimation du quantile est bien inférieure à sa valeur théorique.

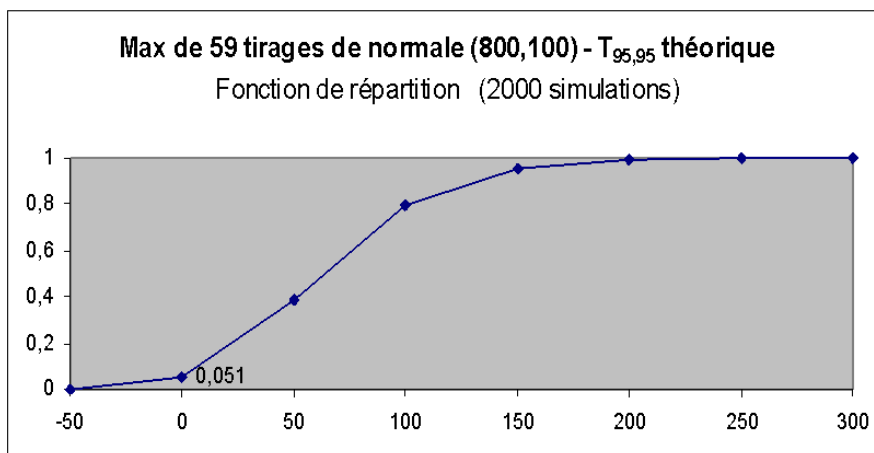


Figure 5. Fonction de répartition de l'écart entre la valeur maximale de l'échantillon et la valeur théorique du quantile (2000 simulations)

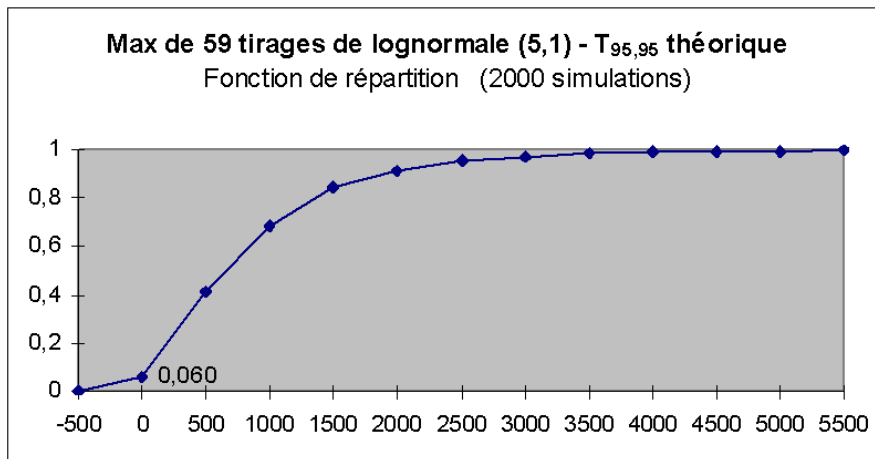


Figure 6. Fonction de répartition de l'écart entre la valeur maximale de l'échantillon et la valeur théorique du quantile (2000 simulations)

2.2. Méthode du Bootstrap

Pour évaluer cette méthode, nous avons, à nouveau, procédé au tirage aléatoire d'un échantillon de 59 valeurs distribuées suivant la loi théorique à partir duquel nous avons généré 250 échantillons « bootstrapés » pour estimer le quantile à 95 % que nous comparons à sa valeur théorique. Réitérant 2000 fois l'opération, nous avons obtenu les fonctions de répartition de l'écart entre les valeurs estimées et théoriques des quantiles présentées aux figures 7 et 8, respectivement pour les lois normale et lognormale. Ces mêmes figures présentent également l'écart obtenu avec la méthode de Wilks (quantile correspondant au maximum des 59 valeurs de l'échantillon avant Bootstrap).

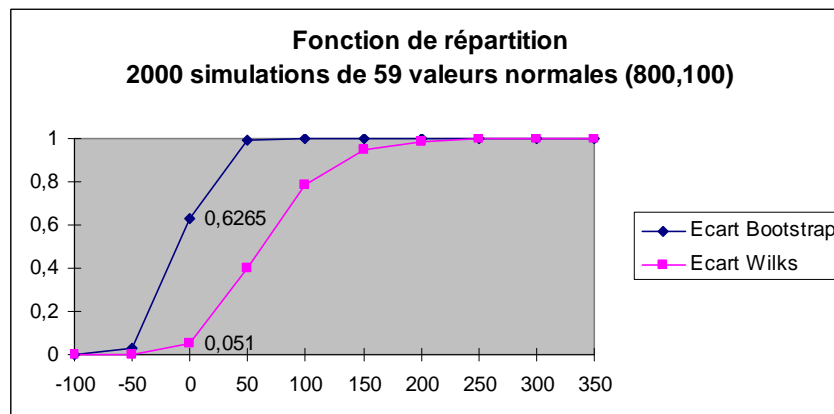


Figure 7. Résultats pour une loi normale

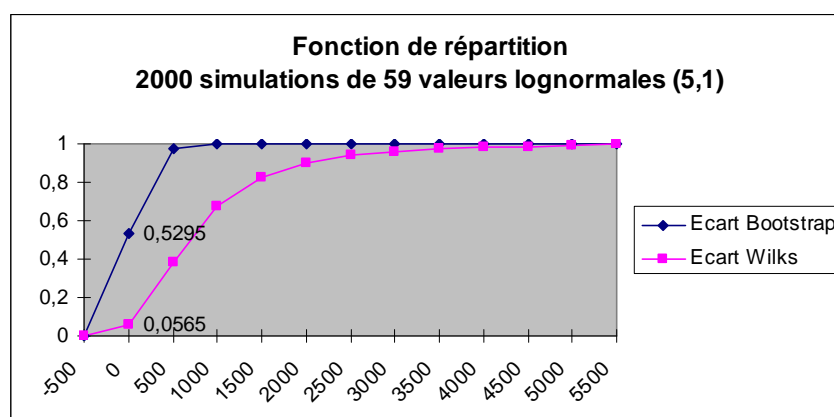


Figure 8. Résultats pour une loi lognormale

La confiance de l'estimateur du quantile à 95 % obtenu par Bootstrap est d'environ 50 %, bien loin des 95 % escomptés. Ce résultat n'est pas vraiment étonnant car l'hypothèse forte considérant que la distribution des quantiles est approximativement normale est d'autant moins vraie que le quantile est en limite de distribution. En effet, les valeurs extrêmes de l'échantillon initial, qui conditionnent l'estimation du quantile, ne sont présentes dans les échantillons générés par Bootstrap que de manière épisodique. Les deux exemples de la figure 9 confirment le fait que cette hypothèse ne peut pas être retenue.

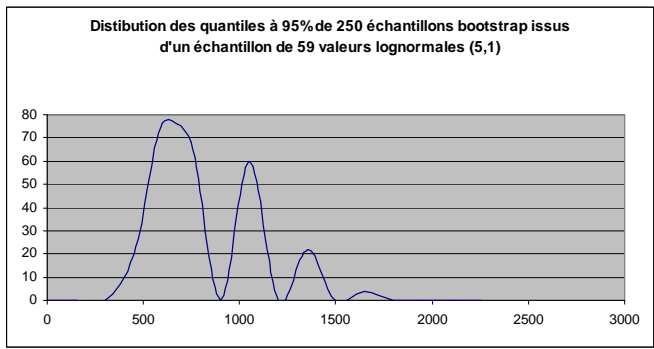
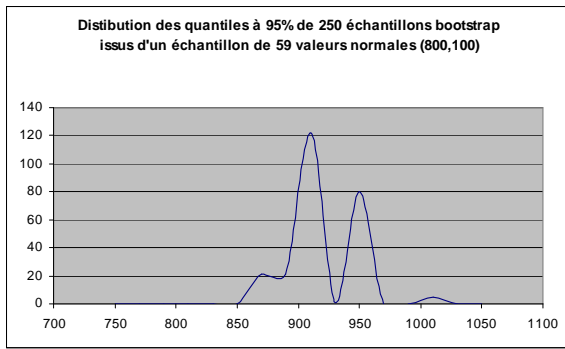


Figure 9. Distribution d'un quantile dans une population « bootstrapée »

Toutefois, les résultats des figures 10 et 11 montrent l'apport (modeste) du bootstrap à l'estimation du quantile à 95 %, de la moyenne et de l'écart-type par rapport au simple traitement effectué directement à partir d'un échantillon (de 59 valeurs). Caractérisé par un resserrement de l'intervalle de confiance, cet apport est réel pour l'estimation de l'écart type mais faible pour celle d'un quantile à 95 % et celle de la moyenne.

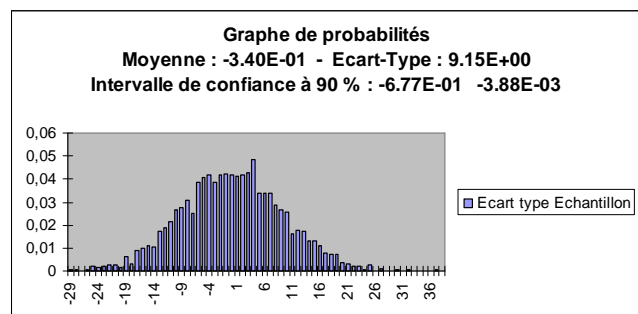
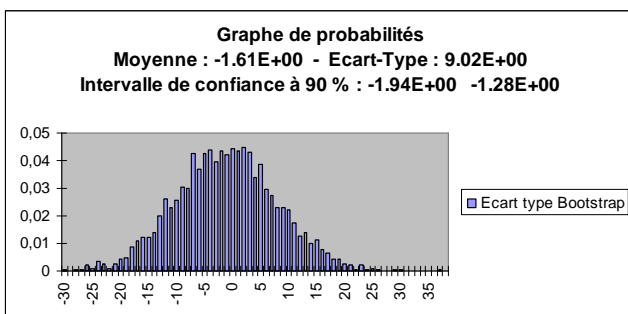
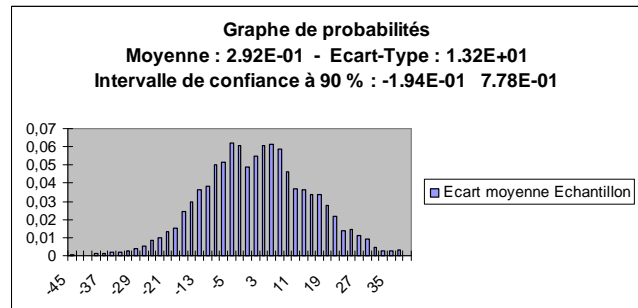
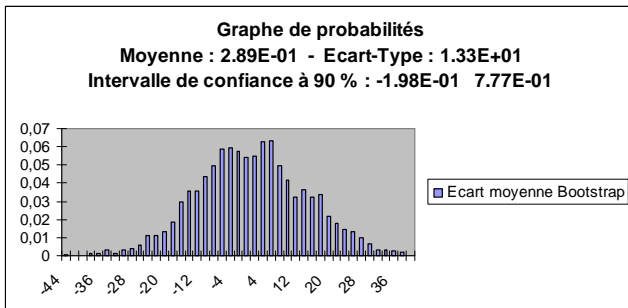
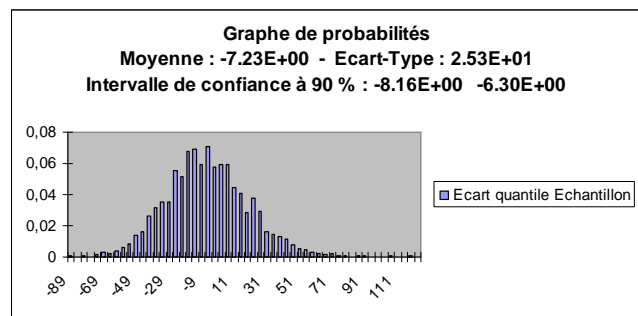
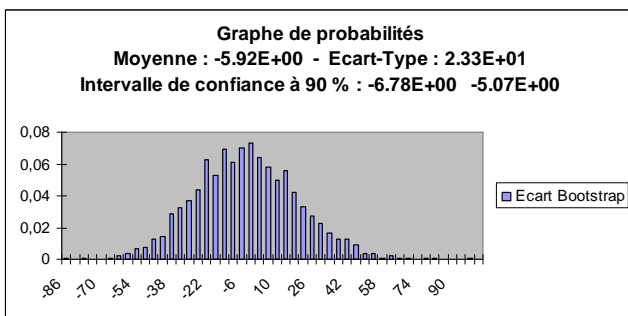


Figure 10. Apport du Bootstrap (loi normale)

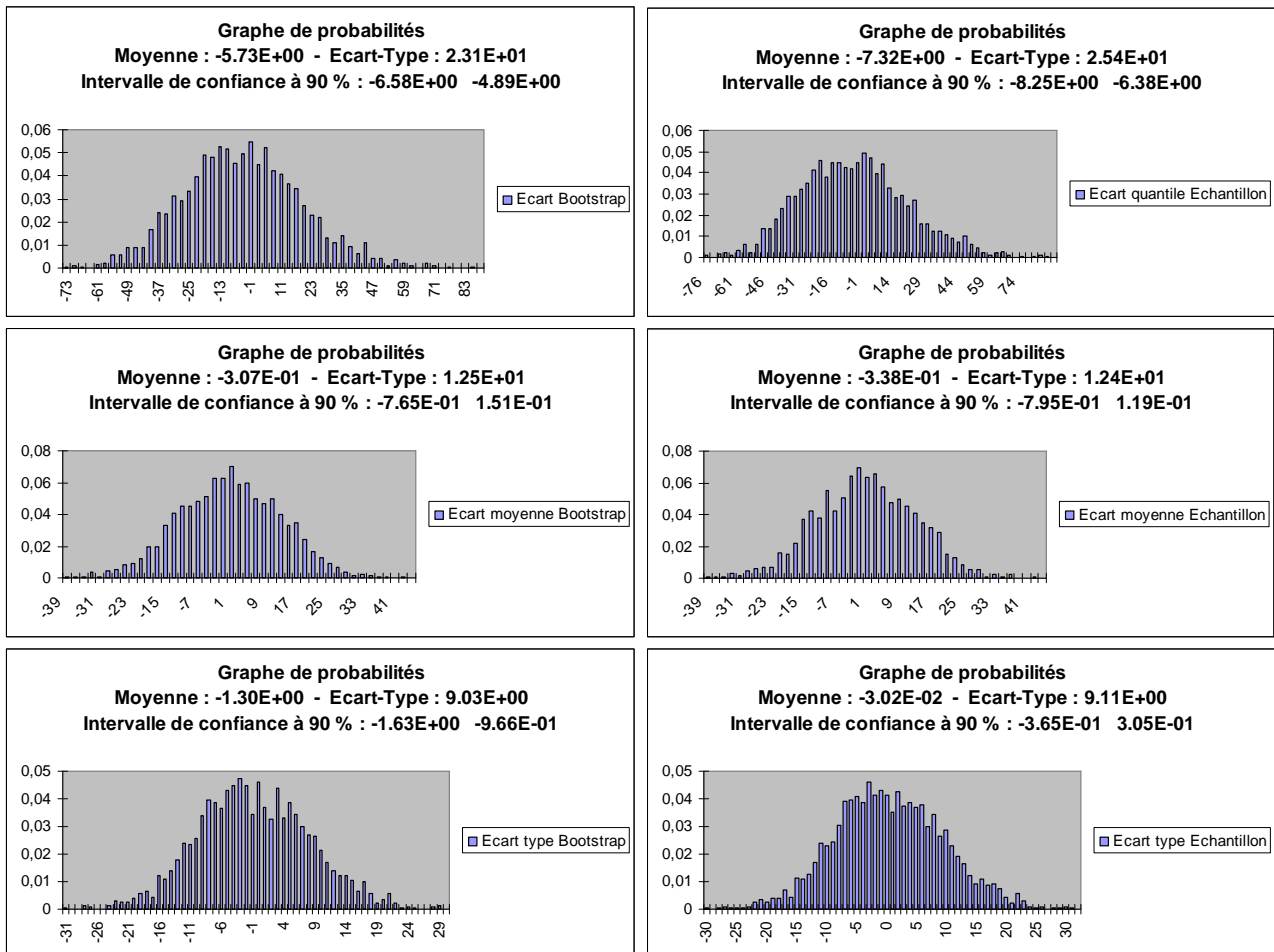


Figure 11. Apport du Bootstrap (loi lognormale)

3. Conclusion

La méthode de Wilks permet de déterminer le nombre minimum de simulations nécessaires à l'obtention d'un majorant de la valeur d'un quantile α au taux de confiance β . Ce nombre est donné par la formule de Wilks qui résulte directement de l'expression de l'intervalle de confiance de la loi binomiale. Aussi n'avons-nous pas été surpris d'obtenir par cette méthode des estimations parfaitement en accord avec la confiance attendue.

Mais il n'en est pas de même pour l'estimation obtenue par la méthode du Bootstrap dont la confiance réelle n'a pas dépassé les 50 % au cours de cette évaluation, bien loin des 95 % escomptés, et dont l'apport s'avère bien modeste par rapport à une estimation réalisée directement à partir de l'échantillon original.

Au terme de cette évaluation, nous considérons que la méthode du Bootstrap doit être rejetée pour estimer un quantile même si son intervalle de confiance peut être très resserré puisque le nombre d'échantillons générés par tirage aléatoire peut être augmenté à loisir. En effet, l'estimateur du quantile obtenu par cette méthode s'avère quelque peu fictif car il ne correspond pas à celui de la population mère mais à celui d'une population virtuelle pouvant être générée à partir d'un échantillon.

Datant de plus d'un demi-siècle, la méthode de Wilks pourrait être utilisée prochainement dans le domaine spatial pour valider des dimensionnements de nature probabiliste à partir de résultats de simulation stochastique. Remplaçant des évaluations déterministes en pire cas, ce type de dimensionnement nous apparaît en effet susceptible de limiter les marges inutiles et de valoriser au mieux les moyens existants.

Références

- [1] WILKS, S. S., « Determination of Sample Sizes for Setting Tolerance Limits », The Annals of Mathematical Statistics, Vol.12, pp. 91-96, 1941.
- [2] EFRON, B., TIBSHIRANI, R. J., An Introduction to the Bootstrap, Chapman and Hall, 1993.
- [3] JOUCLA, J., CHOJNACKI, E., Expertise des méthodes d'évaluation des incertitudes du code CATHARE dans le programme international OCDE BEMUSE, IRSN – Rapport scientifique et technique, 2006.
- [4] CABARBAYE, A - Fichiers de calcul accessibles à l'adresse Internet :<http://perso.wanadoo.fr/andre.cabarbaye/pdf/TP%2017.doc>