

TP SdF N° 17

Estimation d'un quantile par la méthode de Wilks et la méthode du Bootstrap

Ce TP a pour objet d'évaluer deux techniques de maîtrise des incertitudes sur les résultats d'un code de calcul.

Problématique : On cherche à 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 la température critique de 1204°C, dans 95% des cas.

Disposant d'un modèle mathématique du cœur nucléaire, on réalise un nombre restreint de simulations de type Monte-Carlo (chacune durant plusieurs heures), en jouant sur les paramètres du code de calcul, afin d'estimer le quantile α à 95% avec une confiance β elle-même égale à 95% ($T_{\alpha, \beta}$: quantile α au taux de confiance β).

1 – N'ayant aucune connaissance particulière en physique nucléaire, on fait l'hypothèse que la température atteinte peut être distribuée selon une loi normale de moyenne 800°C et d'écart-type 100°C ou de type lognormale de paramètres $\mu = 5$ et $\sigma = 1$.

Calculer pour les deux distributions de température la valeur théorique du quantile à 95 %.

2 – Calculer le nombre de simulations nécessaires à l'obtention du résultat souhaité par application de la méthode de Wilks pour les rang 1, 2 et 3.

Valider cette méthode au rang 1 pour les deux distributions de température, en comparant un grand nombre de fois (>1000) l'estimation obtenue avec les valeurs théoriques du quantile à 95 %.

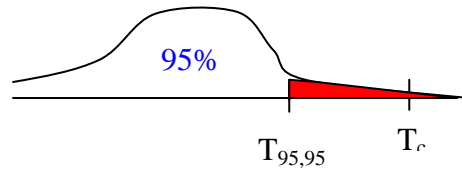
3 – Evaluer ce même quantile par la méthode du Bootstrap au moyen de 250 échantillons générés à partir d'un échantillon original de 59 résultats de simulation.

Chercher à valider cette méthode pour les deux distributions de température, en comparant un grand nombre de fois (>1000) l'estimation obtenue avec les valeurs théoriques du quantile à 95 %.

4 – Conclure sur l'apport de ces méthodes à l'estimation d'un quantile.

1. VALEUR THEORIQUE DU QUANTILE A 95 % DE LA TEMPERATURE.

On note $T_{\alpha, \beta}$ le quantile α que l'on cherche à estimer avec un taux de confiance. Ce quantile doit être inférieur à la température critique ($T_{95, 95} < T_c$).



Dans le cas d'une loi normale de moyenne 800°C et d'écart-type 100°C ou de type lognormale de paramètres $\mu = 5$ et $\sigma = 1$, la valeur théorique du quantile à 95 % de la température se calcule directement sous Excel par les formules suivantes :

$$=LOI.LOGNORMALE.INVERSE(95\%;5;1) = 769^{\circ}\text{C}$$

$$=LOI.NORMAL.EINVERSE(95\%;800;100) = 964^{\circ}\text{C}$$

2. FORMULE DE WILKS¹

La méthode de Wilks est la méthode la plus employée jusqu'à ce jour en thermo-hydraulique nucléaire pour estimer un quantile. Elle permet de déterminer le nombre minimum N de simulations nécessaires pour évaluer un majorant de la valeur recherchée $T_{\alpha, \beta}$.

Ce nombre est donné par la formule de Wilks, qui résulte directement de l'expression de l'intervalle de confiance du paramètre p de la 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 quartile $T_{95,95}$ sera donnée par la valeur pire cas obtenue au cours de N simulations ($r = 1$) avec N donné par l'expression :

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

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

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

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

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

Pour valider cette méthode, on tire aléatoirement un échantillon de 59 valeurs distribuées suivant les lois lognormale ou normale 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 %.

¹ WILKS, S. S., « Determination of Sample Sizes for Setting Tolerance Limits », *The Annals of Mathematical Statistics*, Vol.12, pp. 91-96, 1941.

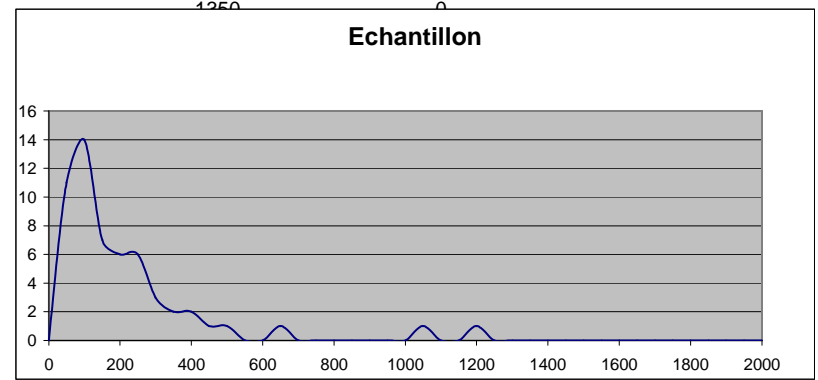
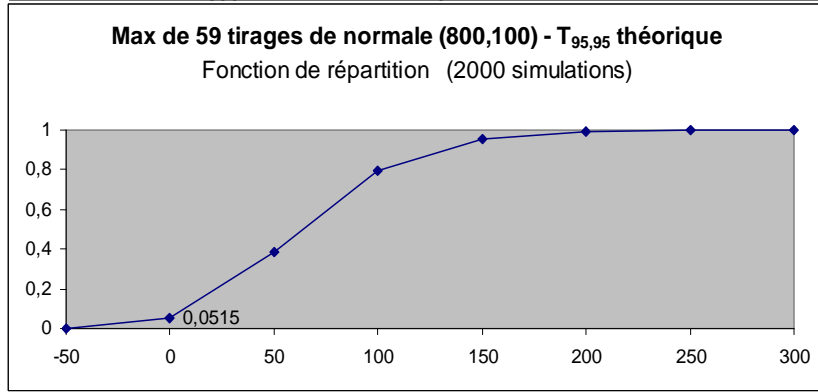
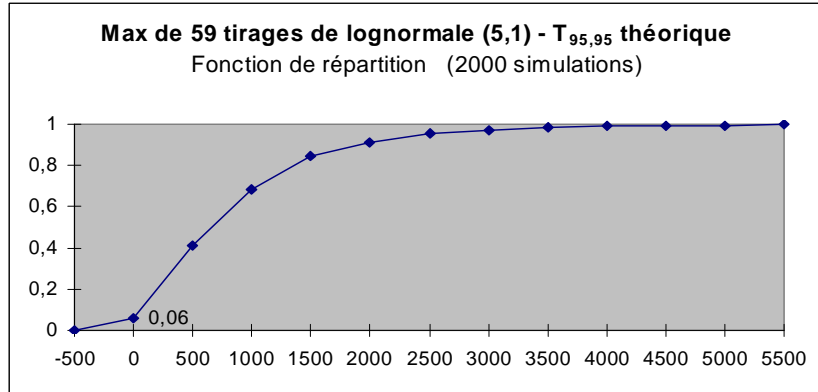
Echantillon

2011,76243
 95,1143697
 101,999918
 35,5196195
 75,3160766
 174,664211
 14,4834687
 76,2237158
 10,8836056
 92,2678544
 79,9832715
 138,787215
 57,8550342
 25,9101532
 179,374911
 230,633555
 164,785929
 1034,00554
 261,236794
 137,77062
 127,835771
 40,643079
 26,2414647
 168,176514
 74,5729451
 347,266501
 68,6856945
 473,106376
 206,05679
 255,575109
 429,780677
 230,504006
 165,69926
 150,284314
 342,855964
 49,8850582
 147,972201
 29,4554673
 3156,12793
 72,3508682
 76,7347538
 18,8374274
 278,40104
 27,1943732
 234,183339
 224,478884
 362,879202
 18,8153342
 61,64516
 90,9053455
 112,609457
 209,00527

Loi : **Lognormale**

	μ	σ
Normale	800	100
Lognormale	5	1

Quantile à :	95%
Théorique :	768,817389
Max Echantillon :	3156,12793
Ecart :	2387,31054



La fonction de répartition de l'écart entre la valeur maximale de l'échantillon et la valeur théorique du quantile, présentée ci-dessus, a été obtenue après 2000 simulations d'un échantillon de 59 valeurs. Dans 95 % des cas environ, l'estimation du quantile est inférieure à sa valeur théorique. Le fichier Excel correspondant est fourni ci-après (ouverture par double clic).



Formule de
 Wilks.xls (27 Ko)

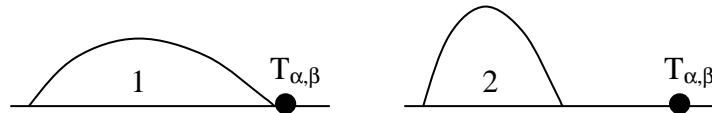
3. METHODE DU BOOTSTRAP²

Proposée à la fin des années 70 par Bradley Efron, le bootstrap est une méthode de suréchantillonnage 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.

Exemple : **2 5 7 8 9**
 ↓
 2 2 7 9 9
 5 8 8 9 9
 2 5 7 8 8
 - - - - -

Même si aucune information nouvelle n'est créée, on cherche à exploiter toute l'information contenue dans l'échantillon original.

A titre d'illustration, les distributions, ci-dessous, représentent 2 échantillons de 59 simulations utilisés pour estimer le quantile $T_{\alpha,\beta}$ par la méthode de Wilks (pire cas). On peut supposer que l'estimation issue de l'échantillon 2 est plus conservative que celle issue de l'échantillon 1.



A partir de l'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 ci-dessous, la moyenne de la population peut être estimée comme étant 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 l'erreur standard du paramètre étudié.

					Moyenne
2	5	7	8	9	6,2
		↓			
2	2	7	9	9	5,8
5	8	8	9	9	7,8
2	5	7	8	8	6
				Moyenne :	6,53
				Ecar type :	1,10

On peut procéder 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.

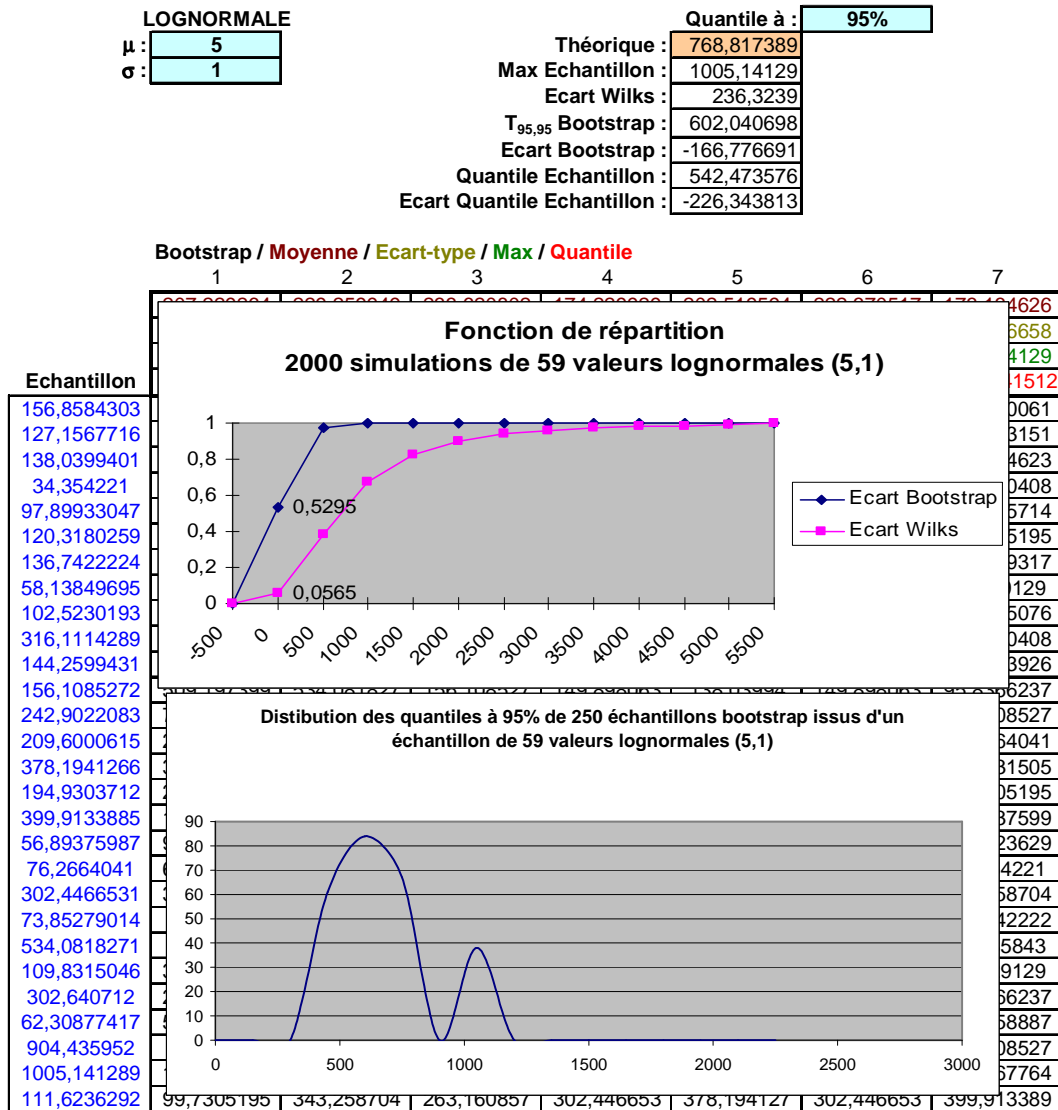
Mais, cette hypothèse est d'autant moins vraie que le quantile est en limite de distribution car les valeurs extrêmes de l'échantillon initial, qui conditionnent alors l'estimation du quantile, ne sont présentes dans les échantillons générés par bootstrap que de manière épisodique.

Par ailleurs, si le suréchantillonnage peut affiner un estimateur par rapport à celui obtenu de manière

² EFRON, B., TIBSHIRANI, R. J., *An Introduction to the Bootstrap*, Chapman and Hall, 1993.

similaire à partir de l'échantillon initial, le quantile à 95% obtenu par bootstrap n'a aucune raison d'être meilleur que celui obtenu par la méthode de Wilks (valeur maximale d'un échantillon de 59 valeurs).

Pour évaluer cette méthode, nous procédons, à nouveau, au tirage aléatoire d'un échantillon de 59 valeurs (distribuées suivant les lois lognormale ou normale) à partir duquel nous générons 250 échantillons bootstrapés pour estimer le quantile à 95 % que nous comparons à sa valeur théorique. Répétant 2000 fois l'opération (avec l'outil SIMCAB), nous obtenons les fonctions de répartition de l'écart entre l'estimation et la valeur thorique suivantes :



Le fichier Excel correspondant est fourni ci-après :



Bootstrap
ognormale.xls (520 .

Loi lognormale (5,1)

