

# OBTENTION D'INTERVALLES DE CONFIANCE EN REASSURANCE PAR LA METHODE DU BOOTSTRAP<sup>1</sup>

Pierre JAL<sup>2</sup>

*Actuarial Consultant*

*Financial & Actuarial Department, Consulting Division*

*Gecalux Group*

## RÉSUMÉ

Cet article propose une méthode d'obtention d'intervalles de confiance pour un taux de réassurance XS, reflétant l'incertitude liée à la faible quantité de données disponible. Nous présenterons d'abord les grands principes du bootstrap en expliquant comment obtenir une estimation de l'erreur standard du paramètre d'intérêt. Nous exposerons ensuite la méthode du bootstrap dite « standard », fondée sur l'hypothèse de normalité de la loi du paramètre d'intérêt ; enfin, nous nous intéresserons plus particulièrement à la méthode des quantiles et à ses diverses améliorations, permettant d'obtenir des intervalles de confiance sous des hypothèses de plus en plus faibles. Nous concluons finalement sur l'amplitude relative des intervalles obtenus.

## ABSTRACT

This paper presents a method allowing to obtain confidence intervals for an XL reinsurance rate regarding the uncertainty related to the lack of available data. We will first explain the main principles of the bootstrap by describing how to obtain an estimation of the standard error for the parameter of interest. Then, we will present the so called « standard method » of the bootstrap, based on the assumption of normality for the parameter of interest's distribution ; finally, we will consider the percentile method and its various upgrades, allowing to obtain confidence intervals under weak and weaker assumptions. In the end, we will conclude on the relative amplitude of those intervals.

## MOTS CLÉS :

Réassurance, intervalles de confiance, bootstrap, cotation, XS.

---

<sup>1</sup> Cet article a été rédigé en 2001 au sein de la société Gerling Konzern Globale, Direction pour la France.

<sup>2</sup> E-Mail : [pierre\\_jal@gecalux.com](mailto:pierre_jal@gecalux.com)

**KEY WORDS**

Reinsurance, confidence intervals, bootstrap, rating, XL.

**1. POSITION DU PROBLÈME**

Le but de cet article est de présenter une méthode permettant d'obtenir des intervalles de confiance pour des taux de réassurance d'une couverture en Excédent de Sinistres. Comme chacun sait, le gros problème des réassureurs est avant tout le peu de données dont ils disposent pour mener à bien leurs études. Ce problème se fait réellement sentir sur la partie qui nous intéressera par la suite, à savoir la cotation d'un XS par la méthode probabiliste. En effet, le but est ici de modéliser les fréquences et les montants des sinistres par des lois de probabilité, puis d'en déduire un taux par un raisonnement de type *Fréquence*  $\times$  *Coût* ; cependant, pour ce faire, il convient d'abord de déterminer ces lois sous jacentes, puis d'en estimer les paramètres. C'est justement là que le bas blesse, du fait du peu de données disponibles, il est clair qu'une certaine imprécision va se retrouver dans l'estimation de ces paramètres ; or justement, les modèles ainsi bâtis y sont extrêmement sensibles. De ce fait, il semble essentiel de chercher à apprécier ce degré d'incertitude afin d'obtenir la meilleure vision possible du risque sous jacent.

**2. PRÉSENTATION DE LA MÉTHODE DU BOOTSTRAP**

Le Bootstrap peut être soit paramétrique, soit non-paramétrique. Dans notre cas, nous ne souhaitons pas déterminer un intervalle de confiance pour une loi totalement connue, mais plutôt, une fois une loi choisie pour modéliser notre variable, déterminer le degré d'imprécision dû à l'erreur d'estimation des paramètres sous jacentes. En conséquence, c'est au Bootstrap « non-paramétrique » que nous aurons recours.

Dans la suite de cet article, nous nous intéresserons seulement à l'incertitude provenant de la loi modélisant les montants des sinistres. En ce qui concerne la modélisation des fréquences de sinistre, nous avons retenu le modèle de Poisson « classique ».

**2.1. Le principe de substitution**

Soit  $(X_1, X_2, \dots, X_n)$  un n-échantillon iid d'une variable aléatoire X de fonction de répartition F et  $(x_1, x_2, \dots, x_n)$  une de ses réalisations.

Soit  $\theta$  un paramètre d'intérêt et  $\hat{\theta}(x_1, x_2, \dots, x_n)$  un estimateur de  $\theta$ .

Soit  $\hat{F}_n$  la fonction de répartition empirique de l'échantillon considéré.

Supposons maintenant que l'on s'intéresse à la variance de  $\hat{\theta}$ , on aura alors le calcul suivant :

$$\left| \begin{aligned} V_F(\hat{\theta}) &= \int [\hat{\theta}(y_1, \dots, y_n) - E_F(\hat{\theta}(y_1, \dots, y_n))]^2 \cdot dF(y_1) \cdot \dots \cdot dF(y_n) \\ E_F(\hat{\theta}(y_1, \dots, y_n)) &= \int \hat{\theta}(y_1, \dots, y_n) \cdot dF(y_1) \cdot \dots \cdot dF(y_n) \end{aligned} \right.$$

Ici, le problème que l'on rencontre vient d'abord du fait que l'on ne connaît pas la loi sous jacente du paramètre  $\theta$ . Utilisant le **théorème de Glivenko**, le principe de substitution nous suggère de remplacer, dans le calcul précédent, la fonction de répartition de l'échantillon par une version lissée de sa fonction de répartition empirique ; en effet, ce théorème énonce que la fonction de répartition empirique d'un échantillon iid converge uniformément vers la fonction de répartition de ce même échantillon.

On obtient ainsi la variance dite « bootstrap » de l'estimateur :

$$V_{BOOT}(\hat{\theta}) = \int [\hat{\theta}(y_1, \dots, y_n) - E_{\hat{F}_n}(\hat{\theta}(y_1, \dots, y_n))]^2 \cdot d\hat{F}_n(y_1) \cdot \dots \cdot d\hat{F}_n(y_n)$$

A ce stade, un autre problème se pose, à savoir qu'en général, il est quasiment impossible d'obtenir  $V_{BOOT}$  de manière analytique car son expression est souvent trop complexe. Afin de contourner ce problème, on a alors recours aux méthodes de Monte-Carlo.

## 2.2. Elaboration d'un échantillon bootstrap et méthode de Monte Carlo

Nous allons dans un premier temps construire un nombre  $B$  d'échantillons de taille  $n$  à partir de l'échantillon de départ. Il s'agit ici de faire un ré-échantillonnage de cet échantillon. Pour ce faire, nous effectuerons des tirages aléatoires avec remise dans l'échantillon initial où tous les éléments se voient attribués la probabilité uniforme  $\frac{1}{n}$ .

A chaque itération, on obtient un échantillon  $(x_1^*, x_2^*, \dots, x_n^*)$  ne contenant que des éléments de l'échantillon de départ. Pour chacun de ces échantillons, on calcule alors une estimation  $\hat{\theta}^*$ . A la fin de la procédure on disposera donc de  $B$  valeurs bootstrap de  $\hat{\theta}$ .

On calcule ensuite :

$$V_{BOOT}^{(B)}(\hat{\theta}) = \frac{1}{B-1} \cdot \sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2$$

où

$$\left| \begin{array}{l} \hat{\theta}^*(b) \text{ correspond à l'estimation de } \theta \text{ pour l'échantillon bootstrap } b. \\ \hat{\theta}^*(\cdot) = \sum_{b=1}^B \frac{\hat{\theta}^*(b)}{B} \end{array} \right.$$

Et d'après la loi des grands nombres,  $V_{BOOT}^{(B)} \xrightarrow{B \rightarrow +\infty} V_{BOOT}$

Ainsi, on approximera l'erreur standard de  $\hat{\theta}$  par la valeur suivante :

$$s.e_{BOOT}^{(B)}(\hat{\theta}) = \sqrt{\sum_{b=1}^B \frac{[\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B-1}}$$

L'erreur standard de  $\hat{\theta}$ , sera donc estimée par l'écart-type de l'échantillon  $(\hat{\theta}^*(1), \dots, \hat{\theta}^*(B))$ . Ici, Efron (EFRON, 1993) suggère de prendre  $B$  égal au moins à 200 afin d'obtenir une estimation convenable.

Ainsi, ce que l'on appelle la méthode du bootstrap est bien la combinaison du principe de substitution et des méthodes de Monte Carlo. Le principe de substitution nous dit que l'on peut remplacer la fonction de répartition théorique de l'échantillon considéré par une version lissée de sa fonction de répartition empirique, les méthodes de Monte Carlo nous permettent ensuite d'obtenir des approximations numériques des expressions ainsi obtenues.

### 3. ECHANTILLON BOOTSTRAP D'UN TAUX DE RÉASSURANCE

Nous allons ici expliquer comment constituer un échantillon Bootstrap pour un taux de réassurance XS. Nous ne rentrerons pas dans les détails de la tarification XS et n'en expliquerons donc que les grandes lignes.

Partant donc d'un échantillon de sinistres répartis sur différentes années historiques, le réassureur cherche à déterminer le taux de prime nécessaire à la couverture d'une tranche XS incluant certaines conditions particulières telles que des reconstitutions ou des franchises aggregate par exemple. Après retraitement des données de manière à obtenir une

statistique « as-if », nous allons déterminer séparément la loi de fréquence annuelle ainsi que la loi de sévérité modélisant le mieux ces données. Les paramètres de ces lois sont estimés à partir de la statistique « as-if » en général par la méthode du maximum de vraisemblance.

Une fois ces lois déterminées, il est alors aisé pour le réassureur de déterminer le taux de prime correspondant au traité considéré en utilisant soit des techniques de simulations, soit des méthodes telles que l'algorithme de Panjer (ou éventuellement des méthodes analytiques en l'absence de clauses particulières). Le but est ici de déterminer l'espérance mathématique de la sinistralité annuelle à charge du traité.

Afin d'obtenir un échantillon Bootstrap de notre taux, nous allons effectuer des rééchantillonnages de notre statistique « as-if » selon la méthode décrite précédemment. Pour chacun de ces nouveaux échantillons, nous allons estimer les paramètres de la loi de sévérité retenue. Attention toutefois, nous ne refaisons pas une inférence statistique complète sur le nouvel échantillon et nous conservons la famille de lois initialement retenue, seuls les paramètres de cette loi de sévérité sont estimés à nouveau. Ainsi, pour chacun de nos nouveaux échantillons de sinistralité, nous disposons d'un couple « loi de fréquence annuelle - loi de sévérité », ce qui nous permet de déterminer totalement notre taux de réassurance en appliquant la méthode décrite ci dessus. Nous obtenons donc un taux de réassurance pour chaque nouvel échantillon, ce qui constitue notre fameux échantillon Bootstrap sur lequel nous travaillerons par la suite.

#### **4. OBTENTION D'INTERVALLES DE CONFIANCE**

Le but de cet article est donc d'obtenir des intervalles de confiance pour le taux d'un traité en Excédent de Sinistre, tenant compte de l'incertitude liée aux erreurs d'estimation des paramètres de la loi modélisant les montants des sinistres. Il existe différentes méthodes utilisant le bootstrap et permettant d'obtenir de tels intervalles, nous ne présenterons que celle qui apparaît comme étant la plus performante et la plus facile à implémenter, à savoir la méthode des quantiles et les différentes améliorations qui lui ont été apportées.

Afin de ne pas alourdir cet article, les démonstrations des différents résultats ne seront présentées qu'en Annexe.

Pour toute la suite, le paramètre d'intérêt  $\hat{\theta}(x_1, x_2, \dots, x_n)$  désignera le taux de réassurance de la tranche considérée pour l'échantillon de sinistres  $(x_1, x_2, \dots, x_n)$ ; afin d'alléger les expressions, nous le noterons par la suite  $\hat{\theta}$ .

#### 4.1. Les données.

Afin d'illustrer notre propos, nous appliquerons nos différents résultats à deux exemples numériques. Nous donnerons pour chacun d'entre eux la statistique as-if correspondante, c'est à dire dont les montants ont déjà été revalorisés.

Ces deux exemples sont développés pour des reconstitutions gratuites et illimitées, la méthode s'applique tout aussi bien pour n'importe quelle condition de reconstitutions ou n'importe quelle clause particulière et ce sans difficultés supplémentaires (et c'est bien là sa force), néanmoins, par soucis de simplification de l'exemple, nous avons préféré conserver cette hypothèse même si elle n'est que très rarement utilisée en pratique pour la réassurance short-tail.

##### ► Exemple 1 :

**EPI : 2.059.110.000**

**Tranche : 2.700.000 XS 300.000**

**Reconstitutions gratuites et illimitées.**

Année	1996	1997	1998	1999	2000	2001
<b>Assiette</b>	2.750.847.474	2.750.847.474	2.526.166.495	2.323.890.131	1.980.522.048	2.083.204.205
<b>Sinistres</b>	2.815.818	1.484.517	2.753.674	3.059.638	1.905.774	5.442.430
	1.454.774	1.255.238	2.485.199	773.541	1.191.108	2.699.445
	708.415	782.635	1.858.911	678.979	1.125.185	1.243.984
	485.728	743.905	1.257.771	600.502	1.064.057	659.382
	348.931	676.892	1.083.030	466.470	590.884	466.494
	348.633	499.262	510.643	421.886	571.733	384.857
	278.163	370.061	506.569	358.208	437.568	343.650
	236.351	346.493	438.932	294.880	415.754	233.247
		346.411	393.353	265.137	397.037	225.957
		287.021	384.223	255.308	326.733	
			333.092	237.671	321.230	
			315.770	237.362	293.105	
			299.972	224.110	263.347	
			251.190		210.313	
			248.994			
			242.182			
			240.622			
			219.735			

## ➤ Exemple 2 :

EPI : 85.000.000

Tranche : 6.000.000 XS 3.000.000

Reconstitutions gratuites et illimitées

Année	1996	1997	1998	1999	2000	2001
Assiette	99.571.493	95.793.458	90.785.604	88.787.436	83.423.324	82.620.000
Sinistres	1.921.620	1.320.905	2.873.123	5.150.155	2.611.447	2.487.968
	1.435.140	997.739	2.191.670	2.646.867	1.394.947	2.254.721
		945.352	1.579.740	1.435.128	995.614	2.099.223
		906.399	1.396.822	907.266	976.576	2.021.474
		797.422	931.594			1.928.175
			916.909			777.490
		902.033				

## 4.2. Intervalles de confiance Bootstrap standards

La première hypothèse que l'on pourrait rationnellement faire serait de supposer que notre paramètre d'intérêt, donc ici le taux de réassurance de la tranche XS considérée, suit une loi Normale d'espérance la « vraie » valeur du paramètre  $\theta$  et d'écart type l'erreur standard telle qu'on l'a estimée précédemment, soit  $\sigma = se_{BOOT}^{(B)}(\hat{\theta})$ .

## ➔ Résultat 1 :

Par un calcul trivial, on peut alors aboutir à un intervalle de confiance symétrique pour  $\theta$  de niveau  $1 - \alpha$  :

$$\left[ \hat{\theta} + z_{\alpha/2} \cdot \sigma; \hat{\theta} - z_{\alpha/2} \cdot \sigma \right]$$

où  $z_{\alpha/2}$  représente le quantile d'ordre  $\alpha/2$  de la loi Normale centrée réduite.

Voyons les résultats obtenus sur nos deux exemples pour une valeur de  $\alpha$  de 10%:

## Exemple 1 :

L'adéquation à une loi de Pareto ( 210.313 ; 1,0548) semble la meilleure, la méthode classique de cotation nous amène au taux de réassurance suivant:

$$\hat{\theta} = 0,2228\%$$

$$z_{\alpha/2} = -1,6449$$

$$\sigma = 0,03439\%$$

On en déduit l'intervalle de confiance à 90% suivant :

$$[0,1662\% ; 0,2793\%]$$

**Exemple 2 :**

L'adéquation à une loi de Pareto ( 777.490 ; 1,4964) semble ici la meilleure, la méthode classique de cotation nous amène au taux de réassurance suivant :

$$\hat{\theta} = 1,7433\%$$

$$z_{\alpha/2} = -1,6449$$

$$\sigma = 1,7794\%$$

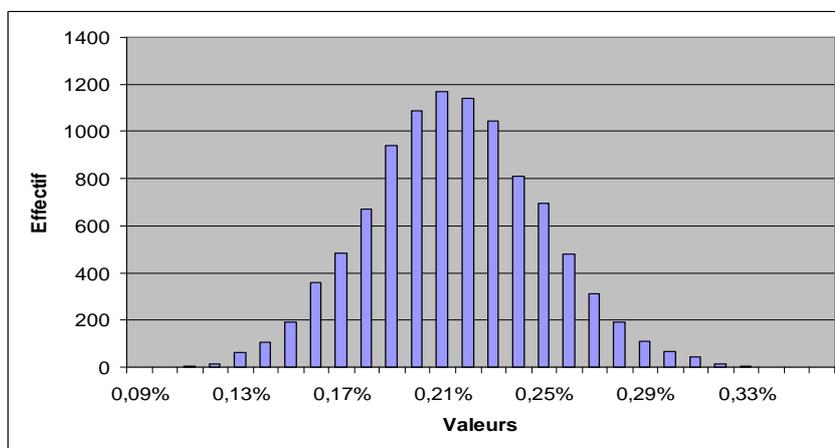
On en déduit l'intervalle de confiance à 90% suivant :

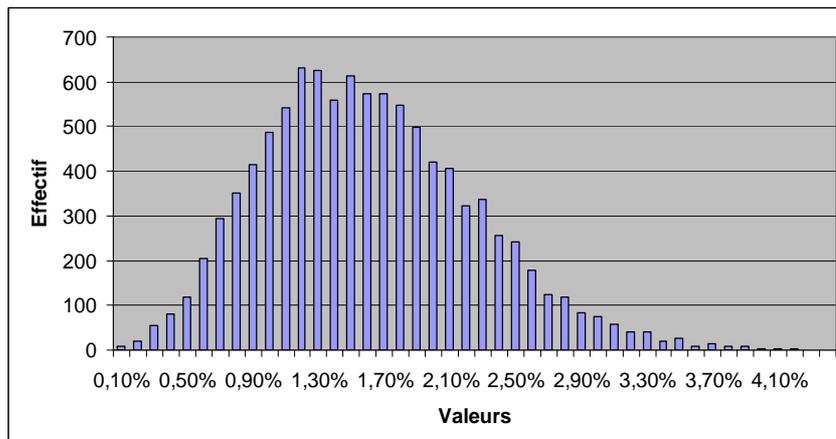
$$[-1,1836\% ; 4,6702\%]$$

Il est ici clair que si l'exemple 1 renvoie des valeurs plausibles, ce n'est manifestement pas le cas du second dont la première borne est négative.

Dans la majorité des cas, les estimateurs des lois de probabilités utilisées pour modéliser les montants des sinistres sont asymptotiquement normaux. Ainsi, lorsque la taille de l'échantillon est suffisamment importante, la distribution de notre paramètre d'intérêt se rapproche de plus en plus d'une distribution gaussienne, ceci peut cependant s'avérer très loin de la réalité lorsque cette « taille critique » n'est pas atteinte.

Ainsi, si l'on représente graphiquement les distributions empiriques des paramètres d'intérêt, on obtient les graphiques suivant :

**Exemple 1 :**

**Exemple 2 :**

Comme le montre clairement l'exemple 2, il semble que l'hypothèse de normalité de  $\hat{\theta}$  soit bien trop forte. En effet, elle ne pourrait être acceptable que pour des échantillons de taille conséquente et sous certaines conditions de régularité ; malheureusement, le monde de la Réassurance qui nous intéresse dans cet article manque cruellement d'échantillons de taille conséquente. La méthode dite des quantiles vise à s'affranchir de cette hypothèse trop restrictive.

**4.3. La méthode des quantiles**

Cette méthode, outre le fait qu'elle ne repose sur aucune hypothèse particulière, a l'avantage d'être extrêmement simple à mettre en œuvre. L'idée ici est de retenir comme intervalle de confiance de notre paramètre d'intérêt un intervalle de confiance empirique du B-échantillon bootstrap de  $\hat{\theta}$ . Ainsi, si l'on note  $F_{BOOT}$  la fonction de répartition empirique bootstrap de  $\hat{\theta}$ , on obtient l'intervalle de confiance de niveau  $1-\alpha$  suivant :

$$\left[ F_{BOOT}^{-1}\left(\frac{\alpha}{2}\right); F_{BOOT}^{-1}\left(1-\frac{\alpha}{2}\right) \right]$$

Par la suite, on montre que lorsque l'hypothèse de normalité du paragraphe précédent est vérifiée, les deux intervalles de confiance obtenus sont identiques. On montre ainsi que sous cette hypothèse, l'intervalle obtenu par la méthode des quantiles est bien le « bon intervalle ». Le réel progrès de cette méthode vient du fait que cet intervalle demeure le « bon » sous l'hypothèse suivante : *il existe une fonction monotone croissante  $f$  telle que  $f(\hat{\theta})$  soit distribuée selon une loi normale d'espérance  $f(\theta)$  et de variance 1.*

Partant de cette hypothèse et en effectuant un calcul analogue à celui de la méthode du bootstrap standard, on aboutit à la formule suivante :

$$P\left(f(\hat{\theta}) + z_{\alpha/2} < f(\theta) < f(\hat{\theta}) - z_{\alpha/2}\right) = 1 - \alpha$$

Dans le cas où  $f$  est connue, il est alors aisé d'obtenir un intervalle de confiance de même niveau pour  $\theta$ . Cependant, en général,  $f$  n'est pas connue, c'est pourquoi on a recours aux méthodes du bootstrap.

**→ Résultat 2 :**

On montre alors que, avec les notations habituelles,

$$P\left(F_{BOOT}^{-1}\left(\frac{\alpha}{2}\right) < \theta < F_{BOOT}^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$

Ainsi, nous avons bien montré que sous l'existence d'une certaine transformation permettant de se ramener à une loi normale, la méthode des quantiles nous donne bien le « bon » intervalle de confiance. L'intérêt de cette remarque tient dans le fait que dans l'expression de l'intervalle de confiance de la méthode des quantiles, il n'est pas nécessaire de connaître l'expression de cette transformation, seule son existence est supposée ! Ceci nous permet donc de travailler avec une hypothèse bien moins contraignante que dans le paragraphe précédent.

En pratique, après avoir généré  $B$  estimateurs bootstrap et les avoir rangé par ordre croissant, on prendra pour bornes de l'intervalle de confiance de niveau  $1 - \alpha$  la  $\left(\left[\frac{\alpha \cdot B}{2}\right] + 1\right)$  ième valeur et la  $\left[\left(1 - \frac{\alpha}{2}\right) \cdot B\right]$  ième valeur (ici les crochets représentent la partie entière).

**Exemple 1 :**

On obtient l'intervalle de confiance à 90% suivant :

$$[0,1636\% ; 0,2774\%]$$

Ici, le gain en précision est assez minime dans la mesure où, comme nous l'avons dit précédemment, la distribution de l'estimateur du taux de réassurance est assez proche d'une loi normale.

**Exemple 2 :**

On obtient ici l'intervalle de confiance à 90% suivant :

$$[0,7085\% ; 2,8102\%]$$

Contrairement à l'exemple précédent, l'intervalle obtenu ici n'a rien à voir avec celui obtenu par la méthode du bootstrap standard. Le gain est réellement important.

#### 4.4. Méthode des quantiles avec correction de biais.

La méthode précédente a permis d'alléger considérablement l'hypothèse nécessaire au modèle ; néanmoins, elle suppose encore que le paramètre, une fois transformé, sera sans biais. En d'autres termes, cela reviendrait à dire que  $P(f(\hat{\theta}) \geq f(\theta)) = 0,5$  ou, comme  $f$  est monotone croissante,  $P(\hat{\theta} \geq \theta) = 0,5$ . Cette hypothèse n'est pourtant que très rarement acceptable, comme le montrent les résultats obtenus sur nos deux exemples :

**Exemple 1 :**  $P(\hat{\theta} \geq \theta) = P(\hat{\theta}^* \geq \hat{\theta}) = 0,5420$

**Exemple 2 :**  $P(\hat{\theta} \geq \theta) = P(\hat{\theta}^* \geq \hat{\theta}) = 0,5833$

Il semble alors nécessaire de tenir compte de ces biais éventuels. La méthode des quantiles avec correction de biais se fonde sur l'hypothèse suivante : *il existe une fonction monotone croissante  $f(\cdot)$  telle que :*

$$\left| \begin{array}{l} f(\hat{\theta}) - f(\theta) \underset{\Delta}{\sim} N(-z_0 \times \sigma ; \sigma^2) \\ f(\hat{\theta}^*) - f(\hat{\theta}) \underset{\Delta}{\sim} N(-z_0 \times \sigma ; \sigma^2) \end{array} \right.$$

Cette hypothèse autorise donc un biais pour la version transformée de notre paramètre d'intérêt. A noter que les deux parties de l'hypothèse n'en font en fait qu'une seule dans la mesure où l'une implique l'autre par application du principe de substitution.

#### → Résultat 3 :

On montre alors que sous cette hypothèse, on obtient l'intervalle de confiance de niveau  $1 - \alpha$  suivant :

$$\left[ F_{BOOT}^{-1} \left( \Phi \left( 2 \cdot z_0 + z_{\alpha/2} \right) \right); F_{BOOT}^{-1} \left( \Phi \left( 2 \cdot z_0 - z_{\alpha/2} \right) \right) \right]$$

avec  $z_0 = \Phi^{-1} \left( F_{BOOT}(\hat{\theta}) \right)$

*Note : On peut remarquer que dans le cas où  $P(\hat{\theta} \leq \theta) = 0,5$ ,  $z_0 = 0$  et on retombe alors sur la méthode des quantiles de base.*

*Note : Insistons une fois de plus sur le fait qu'il n'est pas nécessaire de connaître l'expression de la fonction  $f$ , seule son existence est supposée.*

Appliquons maintenant cette méthode à nos deux exemples :

**Exemple 1 :**

$$z_0 = 0,1054741$$

$$2 \cdot z_0 + z_{\alpha/2} = -1,4340$$

$$2 \cdot z_0 - z_{\alpha/2} = 1,8558$$

$$\Phi\left(2 \cdot z_0 + z_{\alpha/2}\right) = 0,075786$$

$$\Phi\left(2 \cdot z_0 - z_{\alpha/2}\right) = 0,968259$$

On en déduit l'intervalle de confiance à 90% suivant :

$$[0,1705\% ; 0,2851\%]$$

Ici, l'intervalle est modifié de manière assez conséquente, ce qui correspond à la prise en compte du biais que nous avons observé.

**Exemple 2 :**

$$z_0 = 0,2103$$

$$2 \cdot z_0 + z_{\alpha/2} = -1,2242$$

$$2 \cdot z_0 - z_{\alpha/2} = 2,0656$$

$$\Phi\left(2 \cdot z_0 + z_{\alpha/2}\right) = 0,1105$$

$$\Phi\left(2 \cdot z_0 - z_{\alpha/2}\right) = 0,9806$$

On en déduit l'intervalle de confiance à 90% suivant :

$$[0,8747\% ; 3,1674\%]$$

Là encore, l'écart observé avec le résultat de la méthode précédente traduit la prise en compte du biais observé précédemment.

**4.5. Méthode des quantiles avec correction de biais accélérée (BCA) :**

Cette méthode est très proche de la précédente, la différence réside dans l'hypothèse de départ qui est moins restrictive dans le cas de cette méthode que dans la précédente. En effet, on autorise ici l'erreur standard de  $f(\hat{\theta})$  à varier linéairement avec  $f(\theta)$ , ce qui semble être une approximation plus raisonnable de la réalité.

On fait alors l'hypothèse suivante : *il existe une fonction  $f(\cdot)$  strictement monotone, croissante, telle que :*

$$f(\hat{\theta}) - f(\theta) \stackrel{a}{\sim} N\left(-z_0 \times (1 + a \times f(\theta)); (1 + a \times f(\theta))^2\right)^*$$

$$f(\hat{\theta}^*) - f(\hat{\theta}) \stackrel{a}{\sim} N\left(-z_0 \times (1 + a \times f(\hat{\theta})); (1 + a \times f(\hat{\theta}))^2\right)$$

→ **Résultat 4 :**

Partant de cette hypothèse et toujours avec les mêmes notations, on obtient l'intervalle de confiance de niveau  $1-\alpha$  suivant :

$$\left[ F_{BOOT}^{-1} \left( \Phi \left( z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} \right) \right); F_{BOOT}^{-1} \left( \Phi \left( z_0 + \frac{z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})} \right) \right) \right]$$

Comme dans la méthode précédente, on montrera alors que  $z_0 = \Phi^{-1}(F_{BOOT}(\hat{\theta}))$ .

Efron (cf Bibliographie), donne dans son livre une estimation du paramètre  $a$ , obtenue à l'aide du Jackknife :

$$a \approx \frac{\sum_{i=1}^n \frac{(\hat{\theta}^+ - \hat{\theta}_{-i})^3}{6 \cdot \left( \sum_{i=1}^n (\hat{\theta}^+ - \hat{\theta}_{-i})^2 \right)^{1,5}}$$

Où  $\hat{\theta}_{-i}$  représente l'estimateur de  $\theta$  obtenu lorsque l'on retire le  $i$ ème sinistre de l'échantillon original et  $\hat{\theta}^+$  la moyenne de ces estimateurs partiels.

Il est recommandé pour toutes ces méthodes des quantiles de procéder à un nombre de tirages bootstrap plus important que pour déterminer l'erreur standard du paramètre d'intérêt (où l'on recommandait de procéder au moins à 200 tirages). Efron suggère d'effectuer au minimum 1000 tirages pour la détermination d'intervalles de confiance.

**Exemple 1 :**

$$\hat{\theta}^+ = 0,2197\%$$

$$a = 0,01948$$

$$z_0 = 0,1055$$

$$z_{\alpha/2} = -1,6449$$

$$\left| \begin{array}{l} z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} = -1,3891 \\ z_0 + \frac{z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})} = 1,9177 \end{array} \right.$$

$$\left| \begin{array}{l} \Phi \left( z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} \right) = 0,08240 \\ \Phi \left( z_0 + \frac{z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})} \right) = 0,9724 \end{array} \right.$$

On obtient ainsi l'intervalle de confiance à 90% suivant :

$$[0,1722\% ; 0,2873\%]$$

**Exemple 2 :**

$$\left| \begin{array}{l} \hat{\theta}^+ = 1,6846\% \\ a = 0,01707 \\ z_0 = 0,2103 \\ z_{\alpha/2} = -1,6449 \end{array} \right.$$

$$\left| \begin{array}{l} z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} = -1,1900 \\ z_0 + \frac{z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})} = 2,1306 \end{array} \right.$$

$$\left| \Phi \left( z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} \right) = 0,1170 \right.$$

$$\Phi \left( z_0 + \frac{z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})} \right) = 0,9834$$

on obtient ainsi l'intervalle de confiance à 90% suivant :

$$[0,9092\% ; 3,2221\%]$$

## 5. CONCLUSION

Notre but dans cet article n'était pas, nous le rappelons, d'obtenir des intervalles de confiance pour une loi supposée connue après estimation de ses paramètres, auquel cas l'algorithme de Panjer nous aurait permis de résoudre ce problème de manière aisée. Nous avons plutôt cherché à nous intéresser à l'obtention d'intervalles reflétant l'incertitude du taux de réassurance due à la faible quantité d'information disponible (risque d'erreur de spécification de la loi de sévérité). Par son principe de rééchantillonnage, la méthode du bootstrap répond exactement à nos attentes, de plus il nous est possible de l'appliquer à des traités de réassurance incluant toutes sortes de clauses dans la mesure où la méthode de calcul de Monte Carlo permet d'éviter la modélisation analytique du paramètre considéré. Il serait ainsi extrêmement difficile d'obtenir un intervalle de confiance reflétant l'erreur de spécification de manière analytique pour un traité comportant des reconstitutions payantes.

Nous devons cependant préciser que la méthode du Bootstrap ne résout en rien le risque de sous-estimation des tranches hautes. Le Bootstrap doit être manié avec précaution dès que les extrêmes de l'échantillon sont mis en cause, en effet les rééchantillonnages, aussi nombreux soient-ils, ne comporteront jamais de sinistre plus important que le plus grand sinistre historique alors même que la probabilité d'obtenir un sinistre plus important est non nulle. Le réassureur doit conserver cette idée à l'esprit lors de sa tarification.

De plus, la méthode du bootstrap aussi pratique soit elle, n'en est pas magique pour autant ! En conséquence, il ne faut pas être surpris de l'ampleur des intervalles de confiance obtenue ; comme prévu, la faible quantité d'information dont le Réassureur dispose ne lui permet pas de faire des estimations avec une grande précision. Rien ne sert alors d'augmenter le nombre de tirages bootstrap, les résultats ne font que traduire cette incertitude sous-jacente.

Bien que très larges (on peut obtenir des taux allant de plus ou moins 10% à plus ou moins 50%, voir plus dans certains cas extrêmes), ces intervalles de confiance ont l'avantage d'indiquer au souscripteur le degré d'incertitude obtenu, ce qui permettra de

relativiser l'importance du taux renvoyé par les programmes de cotation. Il est alors possible d'établir des guide-lines de souscription plus appropriées à chaque affaire, tenant compte de ces intervalles de confiance.

### BIBLIOGRAPHIE

DERRIG R. A., OSTASZEWSKI K.M. and REMPALA G.A. (2000): « *Applications of Resampling Methods in Actuarial Practise* », Proceedings of the Casualty Actuarial Society, volume LXXXVII

EFRON Bradley (1982): « *The jackknife, the Bootstrap and other resampling plans* », Society for Industrial and Applied Mathematics, Philadelphie

EFRON Bradley (1993): « *An introduction to the Bootstrap* », Chapman & Hall, New York

JAL Pierre (2001): « *Cotation XS en Réassurance Short – Tail* », mémoire de fin d'études ISFA, service reprographie ISFA, Villeurbanne

MANLY Bryan F.J. (1997): « *Randomization, Bootstrap and Monte Carlo methods in Biology* », Chapman & Hall, Londres

### ANNEXE : Démonstrations des résultats

#### → Résultat 1 :

On va chercher  $Z > 0$  tel que : 
$$P\left(\left|\frac{\hat{\theta}-\theta}{\sigma}\right| < Z\right) = 1-\alpha$$

Soit 
$$P\left(-Z < \frac{\hat{\theta}-\theta}{\sigma} < Z\right) = 1-\alpha$$

$$P\left(\frac{\hat{\theta}-\theta}{\sigma} < Z\right) - P\left(\frac{\hat{\theta}-\theta}{\sigma} \leq -Z\right) = 1-\alpha$$

Or par hypothèse,  $\frac{\hat{\theta}-\theta}{\sigma} \rightsquigarrow N(0;1)$ , donc :

$\Phi(Z) - \Phi(-Z) = 1-\alpha$ , où  $\Phi(\cdot)$  représente la fonction de répartition de la loi normale centrée et réduite.

Comme  $\Phi(-Z) = 1 - \Phi(Z)$ , on obtient :

$$1 - 2 \cdot \Phi(-Z) = 1 - \alpha$$

Et donc : 
$$Z = -\Phi^{-1}\left(\frac{\alpha}{2}\right) = -z_{\alpha/2}$$

On obtient donc  $P\left(\hat{\theta} + z_{\alpha/2} \cdot \sigma < \theta < \hat{\theta} - z_{\alpha/2} \cdot \sigma\right) = 1 - \alpha$

D'où l'intervalle de confiance symétrique de niveau  $1 - \alpha$  pour  $\theta$  :

$$\left[ \hat{\theta} + z_{\alpha/2} \cdot \sigma ; \hat{\theta} - z_{\alpha/2} \cdot \sigma \right]$$



→ **Résultat 2 :**

Notons  $G_{BOOT}$  la fonction de répartition empirique bootstrap de  $f(\hat{\theta}^*)$  et  $F_{BOOT}$  celle de  $\hat{\theta}^*$ .

D'après les hypothèses, et par application du principe de substitution, on a :

$$P\left(f(\hat{\theta}^*) + z_{\alpha/2} < f(\hat{\theta}) < f(\hat{\theta}^*) - z_{\alpha/2}\right) = 1 - \alpha$$

Et donc :

$$P\left(f^{-1}\left(f(\hat{\theta}^*) + z_{\alpha/2}\right) < \hat{\theta} < f^{-1}\left(f(\hat{\theta}^*) - z_{\alpha/2}\right)\right) = 1 - \alpha \quad (*)$$

- $G_{BOOT}(f(t)) = P(f(\hat{\theta}^*) < f(t)) = P(\hat{\theta}^* < t) = F_{BOOT}(t)$  car  $f$  est croissante. (\*\*)

- $G_{BOOT}\left(f(\hat{\theta}) + z_{\alpha/2}\right) = P\left(f(\hat{\theta}^*) - f(\hat{\theta}) < z_{\alpha/2}\right) = \frac{\alpha}{2}$

On peut donc écrire :  $f(\hat{\theta}) + z_{\alpha/2} = G_{BOOT}^{-1}\left(\frac{\alpha}{2}\right)$

Et  $f^{-1}\left(f(\hat{\theta}) + z_{\alpha/2}\right) = f^{-1} \cdot G_{BOOT}^{-1}\left(\frac{\alpha}{2}\right) = \left[G_{BOOT}\left(f\left(\frac{\alpha}{2}\right)\right)\right]^{-1}$

d'après (\*\*):  $f^{-1}\left(f(\hat{\theta}) + z_{\alpha/2}\right) = F_{BOOT}^{-1}\left(\frac{\alpha}{2}\right)$  (\*\*\*)

- $G_{BOOT}\left(f(\hat{\theta}) - z_{\alpha/2}\right) = P\left(f(\hat{\theta}^*) - f(\hat{\theta}) < -z_{\alpha/2}\right) = \Phi\left(-z_{\alpha/2}\right) = 1 - \Phi\left(z_{\alpha/2}\right) = 1 - \frac{\alpha}{2}$

On peut donc écrire :  $f(\hat{\theta}) - z_{\alpha/2} = G_{BOOT}^{-1}\left(1 - \frac{\alpha}{2}\right)$

Et  $f^{-1}\left(f(\hat{\theta}) - z_{\alpha/2}\right) = f^{-1} \cdot G_{BOOT}^{-1}\left(1 - \frac{\alpha}{2}\right) = \left[G_{BOOT}\left(f\left(1 - \frac{\alpha}{2}\right)\right)\right]^{-1}$

d'après (\*\*): 
$$f^{-1}\left(f(\hat{\theta}) - z_{\alpha/2}\right) = F_{BOOT}^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (***)$$

- Par application du principe de substitution, on peut réécrire (\*) de la manière suivante :

$$P\left(f^{-1}\left(f(\hat{\theta}) + z_{\alpha/2}\right) < \theta < f^{-1}\left(f(\hat{\theta}) - z_{\alpha/2}\right)\right) = 1 - \alpha$$

En utilisant alors (\*\*\*) et (\*\*\*\*), on obtient :

$$P\left(F_{BOOT}^{-1}\left(\frac{\alpha}{2}\right) < \theta < F_{BOOT}^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$



→ **Résultat 3 :**

- $G_{BOOT}(f(t)) = P(f(\hat{\theta}^*) < f(t)) = P(\hat{\theta}^* < t) = F_{BOOT}(t)$  , car  $f$  est croissante. (\*)

- $P(f(\hat{\theta}^*) \leq f(\hat{\theta})) = P(f(\hat{\theta}^*) - f(\hat{\theta}) \leq 0) = P\left(\frac{f(\hat{\theta}^*) - f(\hat{\theta})}{\sigma} \leq 0\right) = P\left(\frac{f(\hat{\theta}^*) - f(\hat{\theta})}{\sigma} + z_0 \leq z_0\right)$

En utilisant alors l'hypothèse, on obtient :  $P(f(\hat{\theta}^*) \leq f(\hat{\theta})) = \Phi(z_0)$

Or,  $P(f(\hat{\theta}^*) < f(\hat{\theta})) = G_{BOOT}(f(\hat{\theta})) = F_{BOOT}(\hat{\theta})$  , d'après (\*).

On en déduit que  $z_0 = \Phi^{-1}(F_{BOOT}(\hat{\theta}))$

- Par application du principe de substitution, on peut écrire :

$$P\left(f(\hat{\theta}^*) < f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right) = P\left(f(\hat{\theta}) < f(\theta) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right)$$

Et donc :  $G_{BOOT}\left(f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right) = P\left(f(\hat{\theta}) < f(\theta) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right)$

$$G_{BOOT}\left(f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right) = P\left(\frac{f(\hat{\theta}) - f(\theta)}{\sigma} + z_0 < 2 \cdot z_0 \pm z_{\alpha/2}\right)$$

$$G_{BOOT}\left(f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right) = \Phi\left(2 \cdot z_0 \pm z_{\alpha/2}\right)$$

On peut donc écrire :  $f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma = G_{BOOT}^{-1}\left(\Phi\left(2 \cdot z_0 \pm z_{\alpha/2}\right)\right)$  (\*\*)

- Partant de l'hypothèse et selon la méthode habituelle, on montre trivialement que :

$$P\left(f(\theta) \in \left[ f(\hat{\theta}) + z_0 \cdot \sigma + z_{\alpha/2} \cdot \sigma ; f(\hat{\theta}) + z_0 \cdot \sigma - z_{\alpha/2} \cdot \sigma \right]\right) = 1 - \alpha$$

Par croissance de  $f$ , on en déduit :

$$P\left(\theta \in \left[ f^{-1}\left(f(\hat{\theta}) + z_0 \cdot \sigma + z_{\alpha/2} \cdot \sigma\right) ; f^{-1}\left(f(\hat{\theta}) + z_0 \cdot \sigma - z_{\alpha/2} \cdot \sigma\right) \right]\right) = 1 - \alpha$$

Or d'après (\*\*):  $f^{-1}\left(f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right) = f^{-1}\left(G_{BOOT}^{-1}\left(\Phi\left(2 \cdot z_0 \pm z_{\alpha/2}\right)\right)\right)$

$$\begin{aligned} f^{-1}\left(f(\hat{\theta}) + z_0 \cdot \sigma \pm z_{\alpha/2} \cdot \sigma\right) &= (G_{BOOT} \cdot f)^{-1}\left(\Phi\left(2 \cdot z_0 \pm z_{\alpha/2}\right)\right) \\ &= F_{BOOT}^{-1}\left(\Phi\left(2 \cdot z_0 \pm z_{\alpha/2}\right)\right) \quad , \quad \text{d'après (*)}. \end{aligned}$$

D'où le résultat.



→ **Résultat 4 :**

Partant de l'hypothèse et par la méthode habituelle, on montre trivialement que :

$$P\left(z_{\alpha/2} < \frac{f(\hat{\theta}) - f(\theta) + z_0 \cdot (1 + a \cdot f(\theta))}{1 + a \cdot f(\theta)} < -z_{\alpha/2}\right) = 1 - \alpha$$

Ce qui revient à : 
$$P\left(\frac{f(\hat{\theta}) + z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} < f(\theta) < \frac{f(\hat{\theta}) + z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})}\right) = 1 - \alpha$$

Et 
$$P\left(f^{-1}\left(\frac{f(\hat{\theta}) + z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})}\right) < \theta < f^{-1}\left(\frac{f(\hat{\theta}) + z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})}\right)\right) = 1 - \alpha$$

Que l'on notera :  $P(L < \theta < U) = 1 - \alpha$

- Calculons  $L$  :

On peut écrire :

$$F_{BOOT}(L) = P(\hat{\theta}^* < L) = P(f(\hat{\theta}^*) < f(L)) = P\left(\frac{f(\hat{\theta}^*) - f(\hat{\theta})}{1 + a \cdot f(\hat{\theta})} + z_0 < \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} + z_0\right)$$

D'après l'hypothèse, cela revient à écrire :

$$F_{BOOT}(L) = \Phi \left( \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} + z_0 \right) \text{ et donc : } L = F_{BOOT}^{-1} \left( \Phi \left( \frac{z_0 + z_{\alpha/2}}{1 - a \cdot (z_0 + z_{\alpha/2})} + z_0 \right) \right)$$

• Un calcul similaire pour U donne :

$$U = F_{BOOT}^{-1} \left( \Phi \left( \frac{z_0 - z_{\alpha/2}}{1 - a \cdot (z_0 - z_{\alpha/2})} + z_0 \right) \right)$$