LES PROVISIONS TECHNIQUES : UNE APPROCHE PAR SIMULATIONS

Yannick REGAZZONI
Actuaire I.S.F.A.

Jérôme SANDER
Le Foyer Assurance, Luxembourg

Le but de cet article est de proposer une méthode d'estimation des provisions techniques. Contrairement aux méthodes classiques (basées sur les tableaux de développement), nous allons chercher à estimer la loi S du montant total des sinistres.

Pour cela, nous allons construire un n-échantillon de la loi de S à partir de simulations

- de la loi du nombre de sinistres
- de la loi du montant d'un sinistre, éventuellement conditionnée aux paiements déjà effectués

L'échantillon ordonné donne une approximation de la fonction de répartition de S. Par différence avec les montants payés des sinistres, on obtient la loi des provisions techniques.

Cette approche, quoique complexe à mettre en oeuvre, présente de nombreux avantages par rapport aux méthodes classiques :

- Résultats indépendants des cadences de règlements
- Obtention d'intervalles de confiance
- Intégration possible de toute forme de réassurance
- Séparation des provisions pour sinistres tardifs des provisions pour sinistres en cours
- Résultats plus fiables dans les cas suivants :
  - ⇒ fréquence de sinistre faible
  - ⇒ nombre de contrats faible
  - ⇒ dispersion importante des montants de sinistres
  - ⇒ irrégularité dans la cadence de règlement

INTRODUCTION

EXPOSE DU PROBLEME

Les provisions techniques sont définies comme étant la différence à un instant donné entre

- la charge finale (inconnue et donc estimée) de tous les sinistres survenus pendant une période d'observation
- les paiements déjà effectués pour ces mêmes sinistres

Les compagnies d'assurances doivent évaluer ces provisions régulièrement. Il existe pour cela un certain nombre de méthodes, plus ou moins sophistiquées.

BULLETIN FRANÇAIS D'ACTUARIAT, Vol. 1, N° 2, 1997, pp. 81-95

Les procédés d'estimation les plus courants (Chain ladder, London Chain, Cape-Cod, Craighead, Taylor, ...) sont tous basés sur des tableaux de développement, qui ont la forme suivante :

Année d'origine \ ***Année de développement***	0	…	j	…	…	n
0	$C_{0,0}$		$C_{0,j}$			$C_{0,n}$
					…	
i	$C_{i,0}$		$C_{i,j}$			
		…				
n	$C_{n,0}$					

Les Ci,j représentent généralement les montants (cumulés ou non) des paiements.

LES LIMITES DES METHODES CLASSIQUES

La mise en pratique de ces méthodes fait apparaître les insuffisances dont elles souffrent. Les résultats que l'on obtient sont en effet assez médiocres dès que l'on se trouve dans l'une des situations suivantes :

- irrégularité des cadences de règlement (ce qui se produit relativement souvent)
- nombre de contrats trop faible
- fréquence de sinistres trop faible (incendie, grêle, ...)
- dispersion des montants de sinistre élevée (responsabilité civile...)

De plus, il est très difficile, voire impossible

- de séparer les provisions pour sinistres déclarés des provisions pour sinistres tardifs
- d'intégrer la réassurance non proportionnelle, comme les traités en excédent de sinistres
- d'obtenir un intervalle de confiance plutôt qu'une estimation ponctuelle

On notera enfin que les résultats ne sont pas additifs, c'est-à-dire que l'estimation des provisions pour un portefeuille donné n'est pas égale à la somme des estimations pour une partition de ce portefeuille.

VERS UNE AUTRE APPROCHE

Il est possible de contourner ces limites à l'aide d'une autre technique d'évaluation des provisions. Plutôt que de partir d'un historique de développement de paiements, nous allons directement chercher la loi du montant total des sinistres, à partir

- d'une loi du nombre de sinistres (sur une période donnée)
- d'une loi du montant de sinistre

Ces 2 éléments sont la base du modèle, et il est nécessaire de porter une attention toute particulière à leur étude, afin d'en avoir une connaissance la plus exacte possible.

Par simulations informatiques, basées sur nos 2 lois, on génère des réalisations de la variable "Montant total des sinistres". On est donc en mesure de construire un n-échantillon, qui une fois ordonné nous donne une approximation de la fonction de répartition de la loi recherchée.

AVANTAGES DE LA METHODE

- ***Construction d'intervalles de confiance***

L'obtention d'une fonction de répartition nous permet de construire des intervalles de confiance, c'est-à-dire de donner le montant total des sinistres s pour un niveau de risque α donné.

$$s \text{ est tel que } \Pr[\text{Charge sinistre} \leq s] = 1-\alpha$$

- ***Distinction des types de provisions***

La simulation d'une réalisation de la loi du montant total des sinistres passe par la simulation de la charge

⇒ des sinistres en cours
⇒ des sinistres tardifs

On connaît donc directement les provisions liées à ces types de sinistres, lors de la construction du n-échantillon. Rappelons que cette distinction est importante, car les 2 données figurent dans des postes différents au bilan.

- ***Intégration de la réassurance***

Le modèle présenté permet de connaître la loi des provisions brutes de réassurance. Cette information, quoique déjà importante, n'est pas aussi fondamentale que le montant des provisions nettes de réassurance.

Il existe en fait 2 formes de réassurance :

⇒ réassurance proportionnelle (excédent de somme, quote-part)
⇒ réassurance non proportionnelle (excédent de sinistre, excédent de perte annuelle)

Réassurance proportionnelle

Le premier type de traité est tout à fait gérable par les méthodes classiques, puisqu'on a :

*Paiements nets de réassurance = Paiements bruts * coefficient*

Le coefficient est donné par le traité de réassurance. Les données du tableau de développement subissent donc simplement un changement d'échelle.

La méthode par simulation intègre tout aussi bien cette forme de réassurance, puisqu'il suffit de multiplier chaque simulation par ce même coefficient.

Réassurance non proportionnelle

La réassurance non proportionnelle donne une relation du type :

Paiements nets de réassurance = Min(Paiements bruts, Seuil du traité)

Les paiements sont soit considérés sinistre par sinistre (excédent de sinistre), soit cumulés (excédent de perte annuelle).

Les méthodes classiques ne permettent pas l'intégration de cette réassurance. Par contre, notre méthodologie permet, par simple troncature des simulations, d'avoir une vision nette de réassurance. Notons que nous obtenons du même coup une estimation des provisions de la réassurance.

- ***Indépendance de la cadence des règlements***

La méthode n'étant pas basée sur le développement des paiements, aucune des irrégularités ou des changements de cadence n'a d'influence sur le résultat final. En d'autres termes, l'hypothèse de base pour les méthodes classiques (alignement des points entre les années de développement n et n+1, $\forall$ n) n'a pas lieu d'être.

- ***Additivité des résultats***

On peut déterminer une fonction de répartition empirique pour n'importe quel sous-ensemble du portefeuille. Il est tout à fait possible d'additionner plusieurs fonctions de répartition, afin d'obtenir la distribution d'un agrégat plus large.

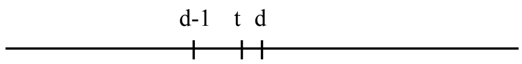
MODELISATION DU PROBLEME

TYPES DE SINISTRES

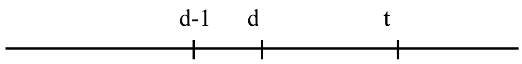
Considérons une période de survenance de sinistres D=[d-1,d] et plaçons nous en date t (date d'observation des données). D est généralement une année civile.

Les situations suivantes peuvent se présenter :

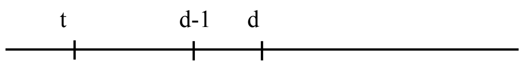
- d-1 < t < d (période en cours)



- d ≤ t (période passée)



- d -1 ≥ t (période future)



Ces graphiques nous permettent d'isoler 3 types de sinistres :

1. **les sinistres déclarés**, c'est-à-dire les sinistres survenus pendant la période d'analyse et déclarés avant t. Cette notion inclut donc les sinistres tardifs connus, c'est-à-dire les sinistres déclarés après d, mais avant t.
2. les **sinistres à venir**, c'est-à-dire les sinistres qui vont survenir et être déclarés dans la période d'analyse. Ce cas n'a de sens que si t < d.
3. les **sinistres tardifs inconnus**, c'est-à-dire les sinistres qui vont survenir pendant la période d'analyse, mais qui seront déclarés après t.

Le modèle que nous allons décrire se veut général, c'est-à-dire capable de traiter tous les cas possibles de d et t. Dans la pratique, les cas les plus courants sont

- $t = d$: on cherche à connaître en fin d'année comptable le montant des provisions pour l'année écoulée
- $t = d + k$ avec k entier ≥ 1 : on cherche à connaître le montant des provisions pour une année de survenance donnée, avec k années de recul.

NOTATIONS

- n^d "Nombre total de **sinistres déclarés**"
- N^v Variable aléatoire "Nombre total de **sinistres à venir**"
- N^t Variable aléatoire "Nombre total de **sinistres tardifs inconnus**"
- N Variable aléatoire "Nombre total de sinistres du portefeuille" = $n^d + N^v + N^t$
- m_j Montant payé au titre du sinistre j vu en t, j=1..N ($m_j = 0$ pour les sinistres à venir et les tardifs)
- S_j Variable aléatoire "Montant du j^ème sinistre, j=1..N "
- S Variable aléatoire "Montant total des sinistres" = $\sum_{j=1}^{N} S_j$
- S_j' Variable aléatoire "Montant du j^ème sinistre sachant que $S_j \geq m_j$, j=1..N " = $S_j / S_j \geq m_j$
- S' Variable aléatoire "Montant total des sinistres connaissant les paiements m_j" = $\sum_{j=1}^{N} S'_j$
- $F_L(x)$ Fonction de répartition de la loi L au point x = $\Pr[L \leq x]$

HYPOTHESES

- Les (S_j , j=1..N) sont indépendants et identiquement distribués
- N est indépendant des (S_j , j=1..N)
- m_j est croissant dans le temps, c'est-à-dire que m_j vu en date a est toujours inférieur ou égal à m_j vu en date b ($\forall$ a $\leq$ b)

REMARQUES

- La variable N est le nombre total de sinistres du portefeuille, **non compris les sinistres sans suite**, c'est-à-dire les sinistres qui sont clôturés sans qu'aucun paiement n'ait été effectué.

Une autre approche est possible, en introduisant une probabilité de classement d'un sinistre en sans-suite.

- Si le sinistre j est un sinistre tardif inconnu ou un sinistre à venir, m_j vaut 0. La loi de $S_j / S_j \geq m_j$ est donc dans ce cas simplement la loi S_j
- Si le sinistre j est un sinistre clôturé, on connaît la réalisation de S_j. m_j est alors le montant final de la charge du sinistre j.

- Il arrive dans la réalité que le montant total des paiements effectués soit décroissant avec le temps. Ceci se produit par exemple lorsque la compagnie exerce un recours contre le responsable du sinistre ou contre l'assuré (franchises, non respect des obligations, ...)

La prise en compte de ces sinistres "négatifs" (c'est-à-dire qui vienne en déduction du montant principal) pourra être traité par l'introduction

⇒ d'une loi sur le nombre de sinistres négatifs
⇒ d'une loi sur le montant d'un sinistre négatif

- $E(S) = \sum_{j=1}^{n^d} E(S'_j) + E(S_j)\,[E(N^t) + E(N^v)]$

PROBLEME

Les provisions techniques valent par définition : $S' - \sum_{j=1}^{n^d} m_j$

Cette expression n'a de sens que si $t \geq d$. En effet, les provisions techniques portent par définition sur les sinistres survenus (déclarés ou non) et non pas sur des sinistres futurs.

Par contre, la variable S' (montant total des sinistres sur la période D) garde toute sa signification, quels que soient d et t. C'est donc sur cette variable que va porter notre étude. On pourra toujours par après en déduire la loi des provisions techniques.

Nous allons donc chercher une fonction F_{Ln} telle que

$F_{Ln}(x) \to F_{S'}(x)$ quand $n \to +\infty$, $\forall\, x$ (convergence presque sûre)

APPROCHE THEORIQUE

IDEE GENERALE

Nous cherchons à connaître le montant final de la charge sinistre. Plutôt que de partir d'un historique de développement de paiements (cf. méthodes classiques), nous allons directement chercher la loi de S'.

L'idée générale de la méthode est de simuler informatiquement un grand nombre de fois des réalisations de la variable S'. Nous obtenons alors un n-échantillon de la loi de S', et donc une idée de sa distribution. La connaissance de S' est d'autant meilleure que n est grand.

La mise en oeuvre de la méthode nécessite de pouvoir simuler des réalisations de

- la loi du nombre de sinistres tardifs inconnus : N^t
- la loi du nombre de sinistres à venir : N^v
- la loi du montant d'un sinistre : S_j

Les simulations de S'_j sont données par conditionnement de S_j sachant que $S_j \geq m_j$.

CONSTRUCTION DE LA FONCTION F_{LN}

Soit (S'^i, i=1..n) une suite de variables aléatoires indépendantes et identiquement distribuées comme S'. On définit F_{Ln} de la façon suivante :

$F_{Ln}(x) = \frac{1}{n}\sum_{i=1}^{n} I(S'^i \leq x)$ avec $I(S'^i \leq x)$ la variable aléatoire indicatrice de l'événement $\{S'^i \leq x\}$

On peut vérifier que F_{Ln} répond bien à notre problème, c'est-à-dire trouver F_{Ln} telle que

$F_{Ln}(x) \to F_{S'}(x)$ quand $n \to +\infty$, $\forall$ x (convergence presque sûre)

En effet, les (S'^i, i=1..n) étant indépendants (par hypothèse), les ($I(S'^i \leq x)$, i=1..n) le sont aussi. D'après la loi forte des grands nombres, on a

$F_{Ln}(x) \to E(I(S'^i \leq x)) = E(I(S' \leq x))$ (convergence presque sûre)

Or $E(I(S' \leq x)) = \int_{-\infty}^{+\infty} I(s' \leq x)\, f_{S'}(x)\, ds'$ où $f_{S'}(x)$ est la fonction de densité de S' au point x

$$= \int_{-\infty}^{x} 1.\, f_{S'}(x)\, ds' + \int_{x}^{+\infty} 0.\, f_{S'}(x)\, ds'$$

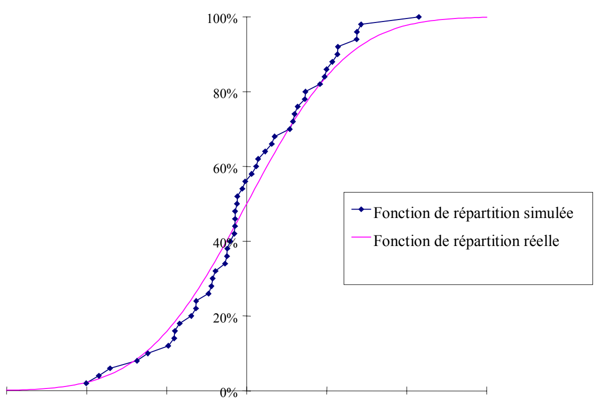
$$= \int_{-\infty}^{x} .\, f_{S'}(x)\, ds'$$

$= F_{S'}(x)$ par définition de la fonction de répartition

On a donc bien

$F_{Ln}(x) \to F_{S'}(x)$ quand $n \to +\infty$, $\forall$ x

Graphiquement, on peut représenter la méthode de la façon suivante :



CAS PARTICULIERS

Dans le cadre de l'analyse d'une période future (t>d), aucun sinistre n'est encore survenu. On a alors

- $n^d = 0$
- $S'_j = S_j \quad \forall j$
- $S' = S$

On peut dans ce cas précis se passer des simulations et arriver directement à un résultat théorique, c'est-à-dire connaître la vraie loi de S.

Cas particulier 1 : Lois additives

S_j est une loi additive si $S^k = \sum_{j=1}^{n} S_k$ est de la même famille de loi que S_j

A titre d'exemple, on pourra avoir

- $S_j \approx \Gamma(\alpha,\beta) \Rightarrow S^k \approx \Gamma(k\alpha, \beta)$
- $S_j \approx N(\mu,\sigma^2) \Rightarrow S^k \approx N(k\mu, k\sigma^2)$

La loi de S est dans ces cas simple à calculer. En effet :

$$\Pr[S \leq s] = E(\Pr[S \leq s / N]) = \sum_{n=1}^{\infty} \Pr[S^n \leq s] \,.\, \Pr[N=n]$$

Cas particulier 2 : convolution de lois

De manière formelle, il est possible de déduire la loi de S à partir de la loi de S_j par convolutions successives : .

$$f_{S^k}(x) = \int f_{S^{k-1}}(x-y).f_{S_j}(y)dy \qquad \text{où } f_L \text{ est la fonction de densité de L}$$

On calcule ainsi successivement les lois de S^2, S^3, ... Ce genre d'approche n'a cependant que peu d'applications pratiques, les équations n'ayant pas de solutions analytiques. Elle n'est en fait possible que pour des lois additives (cf. ci-dessus) et les lois discrètes, c'est-à-dire :

- pour N
- pour S dans les assurances forfaitaires (on paye une indemnité prévue au contrat)

La construction des lois nécessite d'avoir recours à l'outil informatique. Les temps de calculs sont d'autant plus longs qu'il y a de modalités pour les lois N_i et S_j, mais on connaît ici la loi exacte de S.

MISE EN PRATIQUE

SEGMENTATION DU PORTEFEUILLE

Travailler sur un portefeuille de risques hétérogènes n'est pas souhaitable. Il est préférable de constituer des populations homogènes du point de vue du risque (même produit d'assurance, même garantie). Ce découpage est de plus obligatoire, les compagnies d'assurances devant détailler leurs provisions techniques par branche d'assurance.

ESTIMATION DES LOIS DU NOMBRE DE SINISTRES

Nous sommes amenés à estimer la loi N^v (nombre de sinistre à venir) et N^t (nombre de sinistres tardifs). Rappelons que pour ces lois, la base statistique doit exclure les sinistres sans suites, c'est-à-dire les sinistres ne donnant lieu à aucun paiement.

Les lois théoriques possibles pour l'ajustement de ces lois sont entre autres

- la loi Binomiale
- la loi de Poisson
- la loi de Poisson à plusieurs paramètres (Poisson mélange)

ESTIMATION DE LA LOI DU MONTANT DE SINISTRE

En fonction de la durée de la période analysée, il incombe tout d'abord de faire une actualisation des montants de sinistres selon l'inflation.

Les données à extraire sont des montants de sinistres **terminés**. Il ne faut pas ici considérer les sinistres survenus sur une période donnée, car on élimine de fait tous les sinistres longs à traiter, c'est-à-dire les gros sinistres.

Les lois théoriques possibles sont ici

- la loi Normale
- la loi LogNormale (donne généralement de bons résultats)
- la loi Gamma
- la loi LogGamma
- la loi de Pareto et ses dérivées

Remarque importante

Dans la pratique, on trouve fréquemment une loi théorique s'ajustant relativement bien aux données empiriques. Malheureusement, l'ajustement perd souvent de sa qualité dans la queue de distribution. Les différents moments de cette loi (ordre 1, 2, 3...) sont alors différents (voire très différents) des moments empiriques.

Il est donc nécessaire d'ajuster la queue de distribution par une autre courbe, dont les paramètres seront calculés afin d'obtenir

- une continuité de la fonction de répartition théorique
- des moments théoriques proches des moments empiriques

Les fonctions possibles sont multiples (Exponentielle, Pareto, Linéaire, ...)

GENERATEURS DE NOMBRES ALEATOIRES

La création d'un échantillon passe par la mise en place de générateurs de nombres aléatoires. Nous avons besoin pour notre problème de procédures

- pour une loi L quelconque
- pour une loi $L' = L / L \geq a$ (loi conditionnée)

La simulation de nombre aléatoire suivant une loi L donnée peut être effectuée de plusieurs manières

- **à partir de la fonction de répartition F de la loi L**:

si X est une variable aléatoire réelle, de fonction de répartition continue strictement croissante (ce qui est le cas pour une loi de sinistre), on a $F^{-1}(X) \Rightarrow U[0,1]$. On peut donc simuler des nombres aléatoires de la loi L à partir de nombres aléatoires uniformes entre 0 et 1.

- **à partir d'une loi théorique**, dont la fonction de répartition approxime la fonction de répartition empirique :

il existe pour la plupart des lois classiques des algorithmes simples et rapides (voir bibliographie). Cette option présente de nombreux avantages :

⇒ rapidité de traitement : la génération d'un nombre aléatoire d'une loi classique est généralement beaucoup plus rapide que la consultation séquentielle d'un tableau (contenant la fonction de répartition empirique)

⇒ dans les cas de lois additives, on connaît directement la loi de S, sans passer par la simulation

⇒ on dispose d'un modèle plus précis (continuité, queue de distribution)

La simulation de nombres aléatoires suivant une loi conditionnelle $L' = L \,/\, L \geq a$ peut elle aussi être effectuée de plusieurs façons :

- **répétition de simulations jusqu'à réalisation de la condition :**

on simule des réalisations de L jusqu'à ce que la réalisation soit supérieure ou égale à a .

- **à partir de la fonction de répartition inverse**

si on connaît la fonction de répartition inverse, on peut générer des réalisations non plus à partir d'une loi U[0,1], mais une loi U[F(a), 1]. Cette méthode est beaucoup plus rapide pour la loi S'_j, en cas de gros sinistre.

GENERATION D'UN n-ECHANTILLON

La génération d'un n-échantillon de la loi S' passe par la répétition de l'expérience suivante :

$s = 0$
n^s = nombre de sinistres survenus (connu)
n^t = Simulation de la loi N^t
n^v = Simulation de la loi N^v
Pour j=1 à n^s faire
    si le sinistre j n'est pas clôturé ⇒ $s = s +$ simulation de S'_j
    sinon $s = m_j$
Pour j=1 à $n^t + n^v$ faire
    $s = s +$ simulation S_j

Le nombre s obtenu est une réalisation de la loi S'.

CONSTRUCTION DE LA FONCTION DE REPARTITION EMPIRIQUE

On construit f_{Ln} (réalisation de la loi F_{Ln}) à partir de l'échantillon ordonné ($s'_{(1)}$, $s'_{(2)}$, ... $s'_{(n)}$) issu du vecteur aléatoire (S'^{l}, i=1..n) ;

$$f_{Ln}(x) = \frac{k-1}{n} \qquad \text{avec k tel que } s'_{(k-1)} \leq x < s'_{(k)}$$

Pratiquement, on peut améliorer l'approximation en procédant à des interpolations linéaires entre les points $s'_{(k-1)}$ et $s'_{(k)}$ pour k=2..n. On a alors

$$f_{Ln}(x) = \frac{1}{n}\frac{s - s_{k-1}}{s_k - s_{k-1}} + \frac{k-1}{n} \quad \text{avec k tel que } s'_{(k-1)} \leq x < s'_{(k)}$$

EXEMPLE

REMARQUES PRELIMINAIRES

Il n'est pas possible de définir un test statistique commun de la puissance de chaque méthode d'estimation des provisions, mais il est possible de "juger" de leur pertinence à travers une étude historique.

En effet, en appliquant les modèles sur des données passés, on peut vérifier leur qualité par comparaison avec le vrai développement des sinistres, qui aujourd'hui sont terminés et donc connus.

L'exemple (réel) que nous allons traiter est basé sur les année de survenance 1988-1992. Tous les sinistres survenus dans cette période sont aujourd'hui liquidés, et nous connaissons donc le montant final de la charge sinistre.

Notre date d'observation est fin 1992. La notion de sinistre à venir n'a donc pas de sens dans notre exemple.

DONNEES

Paiements cumulés

	Année de développement					Charge finale vue en 1997
Année de survenance	0	1	2	3	4	
1988	5 566 800	9 852 158	10 863 279	10 921 857	10 921 857	10 921 857
1989	10 643 398	18 818 670	18 420 273	18 575 249		18 598 849
1990	5 303 519	7 281 283	18 405 313			19 625 340
1991	9 700 374	16 882 349				17 027 362
1992	8 886 691					13 495 053

On pourra remarquer qu'une forte irrégularité se présente dans le tableau de développement des paiements. Le taux de passage de l'année 0 à l'année 1 est en moyenne de 1.8 à 1.9. Or pour l'année 1990, ce même rapport est supérieur à 3.

Nombre de sinistres

	Année de développement				
Année de survenance	0	1	2	3	4
1988	101	14	3	0	0
1989	91	10	0	2	
1990	124	17	0		
1991	123	15			
1992	129				

Nombre de sinistres terminés (vision fin 1992)

	0	1	2	3	4
88	101	14	3	0	0
89	89	10	0	1	
90	120	17	0		
91	122	13			
92	99				

METHODES CLASSIQUES

Pour les méthodes classiques, nous avons retenu les modèles les plus courants : Chain Ladder et London Chain. Les résultats sont les suivants :

Chain Ladder

	0	1	2	3	4
1988	5 566 800	9 852 158	10 863 279	10 921 857	10 921 857
1989	10 643 398	18 818 670	18 420 273	18 575 249	***18 575 249***
1990	5 303 519	17 281 283	18 405 313	***18 539 536***	***18 539 536***
1991	9 700 374	16 882 349	***17 520 415***	***17 648 185***	***17 648 185***
1992	8 886 691	***17 889 050***	***18 565 164***	***18 700 553***	***18 700 553***

London Chain

	0	1	2	3	4
1988	5 566 800	9 852 158	10 863 279	10 921 857	10 921 857
1989	10 643 398	18 818 670	18 420 273	18 575 249	***18 575 249***
1990	5 303 519	17 281 283	18 405 313	***18 560 098***	***18 560 098***
1991	9 700 374	16 882 349	***17 300 857***	***17 441 554***	***17 441 554***
1992	8 886 691	***16 664 648***	***17 105 470***	***17 243 675***	***17 243 675***

Les estimations pour les années de survenance 1988 à 1991 sont relativement correctes. Par contre, un écart considérable est à relever pour l'année 1992. Cet écart s'explique par la forte irrégularité dans le développement des paiements pour l'année de survenance 1990. Celle-ci va entraîner une surestimation pour l'année de développement 1.

METHODE PAR SIMULATIONS

Regardons ce que donne notre modèle pour cette même année 1992.

Estimation de la loi du nombre de sinistres tardifs

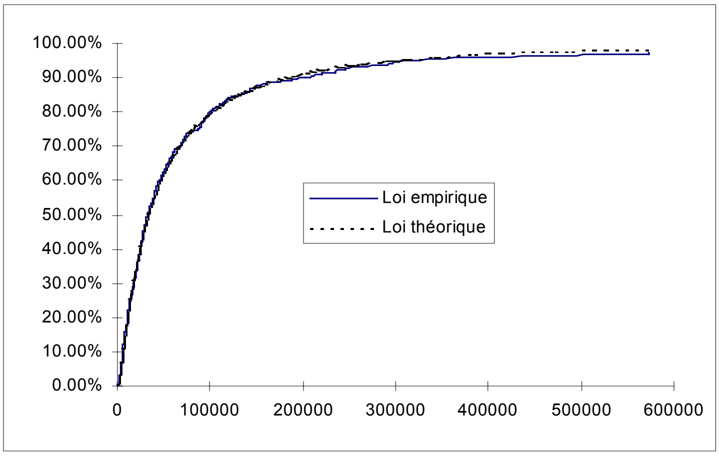
Une rapide étude historique nous montre qu'il y a eu pour les années 1988 à 1991 quelques 58 sinistres tardifs, sur un total cumulé de 3850 garanties. En 1992, le portefeuille est constitué de 1349 contrats. Nous allons retenir pour notre loi une simple loi Binomiale de paramètre n=1349, P=58/3850 = 1.5%.

Estimation de la loi du montant de sinistre

Les sinistres analysés sont ceux dont la date de terminaison est situé entre 1988 et 1992. Les principales statistiques de l'échantillon sont :

- Nombre : 613
- Moyenne : 108 300
- Ecart-type : 403 850

Parmi les différentes lois théoriques testées sur nos données empiriques, c'est un loi logNormale qui semble donnée la meilleure approximation, comme l'illustre le tableau ci-dessous :



Les paramètres de la loi sont

- $\mu = 10.42$
- $\sigma = 1.327$

ce qui donne

- $E(S) = \exp(\mu + \sigma^2/2) = 80\ 859$
- $Var(S) = \exp(2\mu + \sigma^2) . (\exp(\sigma^2) - 1) = 3.15e10 \Rightarrow 177.482$

La mauvaise qualité de l'approximation de la queue de distribution explique les différences observées avec les moments empiriques. Il nous faut donc corriger la loi théorique, avec une autre fonction pour la queue de distribution. Nous allons retenir (arbitrairement) à cette fin une droite d'équation y=ax+b.

$$F_{L'}(x) = \begin{cases} F_L(x) & \text{si } x \leq 300\,000 \\ ax+b & \text{si } 300\,000 < x \leq (1-b)/a \\ 1 & \text{si } x > (1-b)/a \end{cases} \quad \text{où } L \rightarrow LogN(10.42,\ 1.327)$$

Les conditions de continuité et d'égalité entre l'espérance et la moyenne empirique permette de déduire les valeurs de a et b.

- a = 3.8e-8
- b = 0.93927

Estimation de la charge finale

- Les simulations relatives aux sinistres tardifs nous donne en moyenne 2 205 000. Notons que cela correspond bien à ce que l'on pouvait théoriquement calculer : $E(N^t)*E(S) = 1349*1.5\%*108300$
- Les paiements pour les sinistres terminés s'élèvent à 8 836 264 (sur un total de 8 886 691). Les paiements déjà effectués pour les sinistres non terminés sont donc presque nuls.
- Il n'est donc pas surprenant de trouver que les simulations relatives aux sinistres en cours donnent une valeur de 3 250 000, très proche de l'espérance mathématique (nombre de sinistres en cours * coût moyen = 30*108 300)

L'estimation de la charge finale pour l'année 1992 est en conséquence de 14 291 000

- charge pour sinistres tardifs : 2 205 000
- charge des sinistres terminés : 8 836 264
- charge des sinistres en cours : 3 250 000

Conclusions

La vraie charge (vue en 1997) est de 13 495 053. La méthode par simulations donnent une estimation beaucoup plus proche de la réalité (14 291 000) que les deux méthodes classiques analysées.

La qualité de notre estimation repose ici sur le fait qu'elle est indépendante des cadences de règlement. L'irrégularité observée en 1990 entraîne pour toutes les méthodes basées sur les tableaux de développement une surestimation importante de la charge finale pour 1992.

CONCLUSION

Les méthodes d'estimation des provisions techniques basées sur les tableaux de développement ne donnent de bons résultats que lorsque les données réunissent un certain nombre de conditions

- Cadences de règlement régulières
- Nombre important de contrats / sinistres
- Dispersion peu élevé des montants de sinistres

D'autres limitations sont quant à elles liées aux méthodologies :

- estimation ponctuelle plutôt que par intervalles de confiance
- impossibilité d'intégrer certaines formes de réassurance

- impossibilité de distinguer les provisions pour sinistres en cours des provisions pour sinistres tardifs
- non-additivité des résultats

La méthode par simulations est une autre approche du problème. Sa mise en oeuvre reste longue et complexe, car elle nécessite la mise en place d'un outil informatique assez lourd. De plus, elle repose sur la connaissance la plus précise possible de

- la loi du nombre de sinistres
- la loi du montant d'un sinistre

Cependant, en de nombreux aspects, elle se révèle nettement meilleure aux méthodes classiques (indépendance des cadences de règlement, intervalle de confiance, réassurance, additivité des résultats...). Les résultats obtenus sur des données extrêmes (fréquence faible des sinistres, forte dispersion des montants de sinistres...) semblent quant à eux aussi plus corrects, bien qu'aucun test de puissance ne puisse venir démontrer cette affirmation.

BIBLIOGRAPHIE

Christian PARTRAT : Panorama des méthodes statistiques d'évaluation des provisions techniques. Séminaire "Evaluation des provisions techniques en assurances non vie" 1995

G.R. GRIMMET & D.R. STIRZACKER : Probability and random processes. Oxford Science Publications, 1982

G.C. TAYLOR : Claim reserving in non-life insurance. North Holland, 1986

Erwin STRAUB : Non life Insurance Mathematics. Springer Verlag, 1988

Luc DEVRAGE : Non uniform random variate generation

Philippe TASSI : Méthodes statistiques. Economica, 1989

J.L. DOOB : Stochatic Processes. Wiley Classics Library

Donald E. KNUTH : The art of computer programming. Volume 2 : Seminumerical algorithms. Addison-Wesley, 1981

G. CHOW : Econometrics. Mac Graw Hill, 1983