

Assurance dommage

Nicolas Baradel

24 mars 2020

Table des matières

Introduction	2
I Tarification	2
1 Modèles Linéaires Généralisés	3
1.1 Estimation par maximum de vraisemblance	5
1.2 Estimation par maximum de vraisemblance pénalisé	5
2 Tarification a posteriori	7
2.1 Prime Baysienne	8
2.2 Crédibilité non paramétrique	9
II Sinistres tardifs	12
3 Les triangles de liquidation	12
3.1 L'exposition	13
3.2 Risque de réserve	14
3.3 Risque de souscription	14
4 Modèles de Chain Ladder - Mack	14
4.1 Estimation des réserves	14
4.2 Mesure de l'erreur sur l'estimation des réserves	18
4.3 Exemple	19
5 Modèle de Schnieper	21
5.1 Estimation des réserves	21
5.2 Mesure de l'erreur	23
5.3 Le cas particulier du nombre de sinistres au-dessus d'un seuil	24
5.4 Exemple	28
6 Méthode du bootstrap	29
6.1 Modèle de Mack	30
6.2 Modèle de Schnieper	31
7 Prise en compte de l'inflation	32

Introduction

Première partie

Tarifcation

On utilise souvent le modèle collectif en assurance, même pour une tarification individuelle (car dans ce cas, si le nombre de sinistres suit une loi de Bernoulli, on retrouve le modèle individuel). On rappelle le modèle collectif (pour plus de détail, voir [1]), on commence par définir ce qu'est une « somme aléatoire ».

Définition 1. Soient N une variable aléatoire à valeurs dans \mathbb{N} et $(Y_k)_{k \in \mathbb{N}^*}$ une suite de variables aléatoires réelles. La variable aléatoire S définie par :

$$S = \sum_{k=1}^N Y_k$$

est appelée « somme aléatoire ». C'est une somme finie (puisque N est finie), $\forall \omega \in \Omega$,

$$S(\omega) = \sum_{k=1}^{N(\omega)} Y_k(\omega)$$

et $S(\omega) = 0$ si $N(\omega) = 0$. Le terme de « somme aléatoire » vient du fait que, le nombre de Y_k (eux mêmes aléatoires) dans la somme définissant S est aléatoire et dépend de ω .

La variable aléatoire S représente la sinistralité totale, et son écriture en « somme aléatoire » est le modèle collectif. La variable aléatoire N représente le nombre total de sinistre et la suite $(Y_k)_{k \geq 1}$ le montant des sinistres. Si N et Y_1 sont des variables aléatoires de L^1 , alors (voir [1, Première formule de Wald]) :

$$\mathbb{E}(S) = \mathbb{E}(N)\mathbb{E}(Y_1).$$

Dans le cas où N suit une loi de Poisson, on peut également considérer que la sinistralité S^i de chaque individu $i \in \{1, \dots, n\}$ ($n \geq 1$ étant le nombre de contrats) suit son propre modèle collectif (voir [1]). Sa sinistralité s'écrit sous la forme :

$$S^i = \sum_{k=1}^{N^i} Y_k^i.$$

L'espérance du coût des sinistres est ce qu'on appelle la *prime pure* qui est ici : $\mathbb{E}(S^i) = \mathbb{E}(N^i)\mathbb{E}(Y_1^i)$. Afin d'estimer $\mathbb{E}(S^i)$, nous allons estimer $\mathbb{E}(N^i)$ et $\mathbb{E}(Y_1^i)$. Etant donné que chaque individu aura ses caractéristiques propres, il convient d'écrire un modèle permettant d'avoir une prime individuelle. Si la variable aléatoire X^i est à valeurs dans \mathbb{R}^d , nous allons nous intéresser à estimer $\mathbb{E}(N^i | X^i = x^i)$ et $\mathbb{E}(Y_1^i | X^i = x^i)$ avec $x^i \in \mathbb{R}^d$. Nous allons utiliser des modèles linéaires généralisés.

1 Modèles Linéaires Généralisés

Définition 2 (Famille exponentielle). Soit $\mathbb{P}_{\theta \in \mathbb{R}}$ une famille de mesures de probabilité. Le modèle fait parti de la famille exponentielle si cette famille est dominée et que sa densité peut s'écrire sous la forme :

$$f(y; \theta, \phi) = \exp \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right),$$

où a, b et c sont des fonctions mesurables et ϕ est un paramètre de dispersion.

Un premier résultat important est la caractérisation des deux premiers moments.

Lemme 1. Si $Y \sim \mathbb{P}_\theta$ et si b est une fonction C^2 , alors

$$\begin{aligned} \mathbb{E}[Y] &= b'(\theta), \\ \text{Var}(Y) &= b''(\theta)a(\phi). \end{aligned}$$

Démonstration. Si $Y \sim \mathbb{P}_\theta$, on a :

$$\mathbb{E}[\partial_\theta \log(f(Y; \theta, \phi))] = 0.$$

En effet, en notant ν la mesure dominante de la famille :

$$\begin{aligned} \mathbb{E}[\partial_\theta \log(f(Y; \theta, \phi))] &= \mathbb{E} \left[\frac{\partial_\theta f(Y; \theta, \phi)}{f(Y; \theta, \phi)} \right] = \int_{\mathbb{R}} \partial_\theta f(y; \theta, \phi) d\nu(y) \\ &= \partial_\theta \int_{\mathbb{R}} f(y; \theta, \phi) d\nu(y) = \partial_\theta(1) = 0. \end{aligned}$$

On en déduit :

$$\mathbb{E} \left[\frac{Y - b'(\theta)}{a(\phi)} \right] = 0,$$

c'est-à-dire $\mathbb{E}[Y] = b'(\theta)$.

De plus, on a :

$$\mathbb{E}[(\partial_\theta \log(f(Y; \theta, \phi)))^2] = -\mathbb{E}[\partial_\theta^2 \log(f(Y; \theta, \phi))].$$

En effet,

$$\begin{aligned} \mathbb{E}[\partial_\theta^2 \log(f(Y; \theta, \phi))] &= \mathbb{E} \left[\partial_\theta \left(\frac{\partial_\theta f(Y; \theta, \phi)}{f(Y; \theta, \phi)} \right) \right] \\ &= \mathbb{E} \left[\frac{\partial_\theta^2 f(Y; \theta, \phi) f(Y; \theta, \phi) - (\partial_\theta f(Y; \theta, \phi))^2}{f(Y; \theta, \phi)^2} \right] \\ &= \int_{\mathbb{R}} \partial_\theta^2 f(y; \theta, \phi) d\nu(y) - \mathbb{E} \left[\left(\frac{\partial_\theta f(Y; \theta, \phi)}{f(Y; \theta, \phi)} \right)^2 \right] \\ &= \partial_\theta^2 \int_{\mathbb{R}} f(y; \theta, \phi) d\nu(y) = -\mathbb{E}[(\partial_\theta \log(f(Y; \theta, \phi)))^2]. \end{aligned}$$

□

Définition 3 (Fonction de lien canonique). *On suppose que $b' : \mathbb{R} \rightarrow A$ où $A \subset \mathbb{R}$ est bijective. On a $\theta = b'^{-1}(\mu)$. La fonction b'^{-1} est la fonction de lien canonique.*

La plupart des lois habituelles appartiennent à la famille exponentielle.

Exemple 1 (Loi de Bernoulli). *La loi de Bernoulli de paramètre $p \in]0, 1[$ fait parti de la famille exponentielle et sa densité (par rapport à la mesure $\delta_0 + \delta_1$) s'écrit :*

$$f(y; \theta, \phi) = (1 - p)^{1-y} p^y = \exp \left(\frac{y \log \left(\frac{p}{1-p} \right) - \log(1 - p)}{1} + 0 \right).$$

On pose $\theta := \log \left(\frac{p}{1-p} \right)$ qui détermine la fonction de lien canonique. La densité se réécrit :

$$f(y; \theta, \phi) = \exp \left(\frac{y\theta + \log(1 + \exp(\theta))}{1} + 0 \right).$$

On identifie également $b(\theta) = -\log(1 + \exp(\theta))$, $a(\phi) = 1$ et $c(y, \phi) = 0$.

Exemple 2 (Loi de Poisson). *La loi de Poisson de paramètre $\lambda > 0$ fait parti de la famille exponentielle et sa densité (par rapport à la mesure $\sum_{n \in \mathbb{N}} \delta_n$) s'écrit :*

$$f(y; \theta, \phi) = e^{-\lambda} \frac{\lambda^y}{y!} = \exp(y \log(\lambda) - \lambda - \log(y!)).$$

On pose $\theta := \log(\lambda)$ qui détermine la fonction de lien canonique. La densité se réécrit :

$$f(y; \theta, \phi) = \exp(y\theta - \exp(\theta) - \log(y!)).$$

On identifie également $b(\theta) = \exp(\theta)$, $a(\phi) = 1$ et $c(y, \phi) = -\log(y!)$. En pratique on utilise les fonctions de lien $g(\lambda) = \log(\lambda)$ et $g(\lambda) = \sqrt{\lambda}$.

Exemple 3 (Loi Binomiale Négative). *La loi Binomiale Négative de paramètres $r > 0$ et $p \in]0, 1[$ fait parti de la famille exponentielle et sa densité (par rapport à la mesure $\sum_{n \in \mathbb{N}} \delta_n$) s'écrit :*

$$f(y; \theta, \phi) = \frac{\Gamma(r + y)}{y! \Gamma(r)} p^r (1 - p)^y.$$

Son espérance est $\mu := \frac{r(1-p)}{p}$. On paramétrise par (r, μ) , la densité se réécrit :

$$\begin{aligned} f(y; \theta, \phi) &= \frac{\Gamma(r + y)}{y! \Gamma(r)} \left(\frac{r}{r + \mu} \right)^r \left(\frac{\mu}{r + \mu} \right)^y \\ &= \exp \left(y \log \left(\frac{\mu}{r + \mu} \right) + r \log \left(\frac{r}{r + \mu} \right) + \log \left(\frac{\Gamma(r + y)}{y! \Gamma(r)} \right) \right). \end{aligned}$$

On pose $\theta := \frac{\mu}{r + \mu}$ qui détermine la fonction de lien canonique. La densité se réécrit :

$$f(y; \theta, \phi) = \exp \left(y\theta + r \log(1 - \exp(\theta)) + \log \left(\frac{\Gamma(r + y)}{y! \Gamma(r)} \right) \right).$$

On identifie également $b(\theta) = -r \log(1 - \exp(\theta))$, $a(\phi) = 1$ et $c(y, \phi) = \log \left(\frac{\Gamma(r + y)}{y! \Gamma(r)} \right)$. En pratique on utilise les fonctions de lien $g(\mu) = \log(\mu)$ et $g(\mu) = \sqrt{\mu}$.

Exemple 4 (Loi normale).

On suppose observer une variable aléatoire X à valeurs dans \mathbb{R}^d . On suppose de plus qu'il existe une fonction g bijective et $\beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^d$ telle que :

$$g(\mathbb{E}(Y | X)) = \beta_0 + X'\beta.$$

La fonction g est appelée *fonction de lien*. Il est possible d'utiliser la *fonction de lien canonique* du modèle.

La quantité $\theta(X) := \beta_0 + \beta X$ est appelée *le prédicteur linéaire*. Il vient :

$$\mathbb{E}(Y | X) = g^{-1}(\beta_0 + X'\beta) := \mu(X).$$

La difficulté est d'estimer $(\beta_0, \beta) \in \mathbb{R}^{d+1}$.

1.1 Estimation par maximum de vraisemblance

On suppose que $Y | X$ suit la loi $\mathbb{P}_{\theta(X)}$. En particulier, si on observe $(y_i, x_i)_{1 \leq i \leq n}$ variables aléatoires i.i.d., la vraisemblance s'écrit :

$$\ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}) = \sum_{i=1}^n \log(f(y_i; \beta_0 + x_i'\beta, \phi)).$$

Exemple 5 (Loi de Bernoulli). *On obtient la log-vraisemblance :*

$$\ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}) = \sum_{i=1}^n \log \left[\left(\frac{e^{\beta_0 + x_i'\beta}}{1 + e^{\beta_0 + x_i'\beta}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i'\beta}} \right)^{1-y_i} \right].$$

Exemple 6 (Loi de Poisson).

Exemple 7 (Loi normale).

Les équations du premier ordre du problème :

$$(\hat{\beta}_0, \hat{\beta}) \in \arg \max_{(\beta_0, \beta) \in \mathbb{R}^{d+1}} \ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n})$$

ne permettent pas d'en déduire, en général, une solution explicite (en dehors de cas particulier, comme le modèle gaussien avec fonction de lien canonique qui permet de retrouver le modèle linéaire avec bruit gaussien). Nous passons par des méthodes numériques. Dans R, la fonction `glm` permet d'estimer des modèles linéaires. Il convient ensuite d'éliminer les facteurs non déterminants via des tests d'hypothèse.

1.2 Estimation par maximum de vraisemblance pénalisé

Il est possible d'intégrer l'élimination des facteurs non déterminants au modèle en utilisant des méthodes de pénalisation. Soit $\lambda > 0$, on introduit :

$$\ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}; \lambda) := \ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}) - \lambda \|\beta\|_p^p,$$

où $\|\cdot\|_p$ est la norme $p \in \{1, 2\}$ et λ est le facteur de pénalisation. Afin que le modèle soit invariant par facteur d'échelle dans l'estimation des paramètres, il est fondamental de centrer

et réduire les variables explicatives non catégorielles x au préalable. Dans la maximisation, il y a un arbitrage entre l'apport dans la log vraisemblance et la pénalisation.

Lorsque $p = 1$, on parle de pénalisation **Lasso** et lorsque $p = 2$ on parle de pénalisation **Ridge**. À $\lambda > 0$ fixé, on résout le problème :

$$(\hat{\beta}_0, \hat{\beta}) \in \arg \max_{(\beta_0, \beta) \in \mathbb{R}^{d+1}} \ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}; \lambda).$$

La difficulté est alors de choisir un bon $\lambda > 0$. On procède par validation croisée. On découpe l'échantillon de manière aléatoire en $\kappa \geq 2$ parties de taille égale ou semblable. On peut choisir $\kappa = n$ le nombre d'observations, ce qui revient à découper individuellement. On note \mathcal{P}_k la partie $k \in \{1, \dots, \kappa\}$ et l'ensemble forme une partition de $\{1, \dots, n\}$, c'est à dire, on a la réunion disjointe :

$$\bigcup_{k=1}^{\kappa} \mathcal{P}_k = \{1, \dots, n\}.$$

Définition 4. (*Validation croisée*) Soit $k \in \{1, \dots, \kappa\}$ et $\lambda > 0$. L'estimation par validation croisée consiste à estimer le modèle sur l'échantillon observé $(y_i, x_i)_{i \notin \mathcal{P}_k}$. On en déduit $(\hat{\beta}_0^{\lambda, k}, \hat{\beta}^{\lambda, k})$. On calcul ensuite la vraisemblance, avec les paramètres estimés, sur l'échantillon de test :

$$\mathfrak{l}_k(\lambda) := \sum_{i \in \mathcal{P}_k} \log \left(f(y_i; \hat{\beta}_0^{\lambda, k} + x'_i \hat{\beta}^{\lambda, k}, \phi) \right).$$

On introduit :

$$\mathfrak{l}(\lambda) := \sum_{k=1}^{\kappa} \mathfrak{l}_k(\lambda).$$

Le modèle sélectionné est celui qui résout le problème :

$$\hat{\lambda} \in \arg \max_{\lambda > 0} \mathfrak{l}(\lambda).$$

Une fois $\hat{\lambda}$ déterminé, on estime les paramètres sur l'ensemble des données :

$$(\hat{\beta}_0, \hat{\beta}) \in \arg \max_{(\beta_0, \beta) \in \mathbb{R}^{d+1}} \ell(\beta_0, \beta; (y_i | x_i)_{1 \leq i \leq n}; \hat{\lambda}).$$

L'idéal est d'avoir une partition très fine : on estime sur un grand sous échantillon pour tester sur un petit échantillon. Mais plus la partition est fine, plus le temps de calcul est élevé, il y a souvent un arbitrage à faire. En pratique, le minimum étant d'avoir au moins les deux tiers de l'échantillon afin d'estimer les paramètres, et l'idéal étant d'avoir tout l'échantillon sauf un point qui servira de test.

La pénalisation permet d'avoir un modèle bien posé même si le nombre de variables explicatives dépasse le nombre d'observations. Dans ce cas, beaucoup de paramètres seront proches de 0 (pénalisation **Ridge**) ou égaux à 0 (pénalisation **Lasso**).

On note $(\hat{\beta}_0^N, \hat{\beta}^N)$ les paramètres estimés du modèle pour le nombre de sinistres et g_N sa fonction de lien. On note $(\hat{\beta}_0^Y, \hat{\beta}^Y)$ les paramètres estimés du modèle pour le nombre de sinistres et g_Y sa fonction de lien. La prime pure estimée pour l'assuré i de variables explicatives x_i , notée \hat{P}_i , est :

$$\hat{P}_i := g_N^{-1}(\hat{\beta}_0^N + x'_i \hat{\beta}^N) g_Y^{-1}(\hat{\beta}_0^Y + x'_i \hat{\beta}^Y).$$

2 Tarification a posteriori

Malgré la différenciation des individus à travers leurs variables explicatives $x_i \in \mathbb{R}^d$, ceux-ci ont des caractéristiques propres inobservées. Ainsi, même si l'hypothèse que le nombre de sinistres de chaque individu survient à travers une loi de Poisson d'un paramètre qui lui est propre semble raisonnable, on observe souvent une variance supérieure à la moyenne. Une telle propriété apparaît naturellement en conséquence de caractéristiques individuelles inobservables. En effet, supposons que :

$$Y_i \mid \{X_i = x_i, \Theta_i = \theta_i\} \stackrel{i.i.d.}{\sim} \mathcal{P}(\lambda(x_i, \theta_i))$$

Mais que nous ne connaissons que x_i , la caractéristique θ_i est inconnue et propre à l'assuré. Dans ce cas son espérance est :

$$\mathbb{E}[Y_i \mid X_i = x_i] = \mathbb{E}[\mathbb{E}(Y_i \mid X_i = x_i, \Theta)] = \mathbb{E}[\lambda(x_i, \Theta)].$$

Tandis que, par la formule de la variance totale, on a :

$$\begin{aligned} \text{Var}(Y_i \mid X_i = x_i) &= \mathbb{E}[\text{Var}(Y_i \mid X_i = x_i, \Theta_i)] + \text{Var}[\mathbb{E}(Y_i \mid X_i = x_i, \Theta_i)] \\ &= \mathbb{E}[\lambda(x_i, \Theta_i)] + \text{Var}[\lambda(x_i, \Theta_i)] \end{aligned}$$

Si $\lambda(x_i, \Theta_i)$ n'est pas une variable aléatoire p.s. constante, on a $\text{Var}(Y_i \mid X_i = x_i) > \mathbb{E}[Y_i \mid X_i = x_i]$. Un cas intéressant est lorsque $\lambda(x_i, \Theta_i)$ suit une loi Gamma. Dans ce cas la loi du nombre de sinistres de l'assuré pris individuellement suit une loi binomiale négative dont la variance est supérieure à la moyenne.

Lemme 2. Soit $\Theta \sim \mathcal{G}(\alpha, \beta)$ et $Y \mid \{\Theta = \theta\} \sim \mathcal{P}(\theta)$ avec $\alpha > 0, \beta > 0, \theta > 0$. Alors Y suit une loi Binomiale Négative de paramètres

$$\begin{aligned} r &= \alpha, \\ p &= \frac{\beta}{1 + \beta}. \end{aligned}$$

Démonstration. Soit $n \in \mathbb{N}$. On a :

$$\begin{aligned} \mathbb{P}(Y = n) &= \mathbb{E}[\mathbb{E}(\mathbf{1}_{\{n\}}(Y) \mid \Theta)] \\ &= \mathbb{E}\left[\exp(-\Theta) \frac{\Theta^n}{n!}\right] \\ &= \int_{\mathbb{R}_+} \exp(-\theta) \frac{\theta^n}{n!} \frac{\beta^\alpha}{\Gamma(\alpha)} \exp(-\beta\theta) \theta^{\alpha-1} d\theta \\ &= \frac{\beta^\alpha}{n! \Gamma(\alpha)} \int_{\mathbb{R}_+} \exp(-(1 + \beta)\theta) \theta^{n+\alpha-1} d\theta \\ &= \frac{\beta^\alpha}{n! \Gamma(\alpha) (1 + \beta)^{n+\alpha}} \int_{\mathbb{R}_+} \exp(-t) t^{n+\alpha-1} dt \\ &= \frac{\Gamma(\alpha + n)}{n! \Gamma(\alpha)} \left(\frac{\beta}{1 + \beta}\right)^\alpha \left(\frac{1}{1 + \beta}\right)^n \end{aligned}$$

□

Toutefois, l'assureur va pouvoir observer au fil de l'eau le nombre de sinistres de l'assuré et leurs coûts. Cela va lui permettre d'obtenir de l'information sur la variable aléatoire inconnue Θ et sa sinistralité réelle.

2.1 Prime Baysienne

On suppose avoir à notre disposition la prime pure annuelle a priori de chaque assuré i :

$$\hat{P}_i := g_N^{-1}(\hat{\beta}_0^N + x'_i \hat{\beta}_0^N) g_Y^{-1}(\hat{\beta}_0^Y + x'_i \hat{\beta}_0^Y) := \hat{P}_i^N \hat{P}_i^Y.$$

Toutefois, chaque assuré aura des caractéristiques individuelles inconnues. On suppose que son nombre de sinistres N_t sur l'intervalle de temps $[0, t]$ avec $t > 0$ suit une loi de Poisson de paramètre Θt avec $\Theta \sim \mathcal{G}(\alpha_0, \beta_0)$ tel que

$$\frac{\alpha_0}{\beta_0} = \hat{P}_i^N.$$

Sur l'intervalle $[0, 1]$, $\mathbb{E}(N_1) = \mathbb{E}[\mathbb{E}(N_1 | \Theta)] = \mathbb{E}[\Theta] = \frac{\alpha_0}{\beta_0}$ qui est consistant avec la définition de la prime pure.

Lemme 3. *On a :*

$$\Theta | \{N_t = n\} \sim \mathcal{G}(\alpha_0 + n, \beta_0 + t).$$

Démonstration. Il s'agit d'un calcul de loi a posteriori. La loi $\Theta | \{N_t = n\}$ est par construction absolument continue par rapport à la mesure de Lebesgue, pour $\theta > 0$, on a presque partout :

$$\begin{aligned} f(\theta | N_t = n) &\propto f(n | \Theta = \theta) f(\theta) \\ &\propto e^{-\theta t} \frac{(\theta t)^n}{n!} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} e^{-\beta_0 \theta} \theta^{\alpha_0 - 1} \\ &\propto e^{-(\beta_0 + t)\theta} \theta^{\alpha_0 + n - 1}. \end{aligned}$$

On reconnaît la loi $\mathcal{G}(\alpha_0 + n, \beta_0 + t)$. □

Corollaire 1. *La prime pure conditionnellement à l'observation $\{N_t = n\}$ est :*

$$\tilde{P}_i := \mathbb{E}(N_{t+1} - N_t | N_t = n) \hat{P}_i^Y = \left[(1 - z_t) \hat{P}_i^N + z_t \frac{n}{t} \right] \hat{P}_i^Y.$$

avec :

$$z_t := \frac{t}{\beta_0 + t}, \quad t \geq 0.$$

Démonstration. On a :

$$\mathbb{E}(N_{t+1} - N_t | N_t = n) = \frac{\alpha_0 + n}{\beta_0 + t} = \frac{\beta_0}{\beta_0 + t} \frac{\alpha_0}{\beta_0} + \frac{t}{\beta_0 + t} \frac{n}{t},$$

et on en déduit le résultat. □

2.2 Crédibilité non paramétrique

Le cadre bayésien exige une hypothèse de loi sur le paramètre inconnu. La prime qui était obtenue précédemment correspondait à l'espérance conditionnelle, par exemple pour le nombre de sinistres si ce dernier suit un processus $(N_t)_{t \geq 0}$, et si en $t \geq 0$, on observe $\{N_t = n\}$, la prime pure était unitaire par sinistre était par construction :

$$\mathbb{E}(N_{t+1} - N_t \mid N_t = n).$$

Ici, nous allons prendre une forme linéaire en l'observation N_t , ce qui permettra d'aboutir à une prime unitaire de la forme :

$$\left[(1 - z)\hat{P}_i^N + z\frac{n}{t} \right].$$

Nous travaillerons directement sur le coût total des sinistres, mais ce qui suit pourra s'appliquer sur le nombre de sinistres.

Hypothèse 1. Soit $(S_t^i)_{t \geq 0}$ le processus du coût total des sinistres de l'individu $i \geq 1$ et $x_i \in \mathbb{R}^d$ ses caractéristiques. On suppose que, à $i \geq 1$ fixé, la loi de $(S_{t+1}^i - S_t^i) \mid \{X_i = x_i, \Theta_i = \theta_i\}$ est i.i.d. pour tout $t \in \mathbb{N}$.

On peut remplacer le processus du coût total des sinistres par celui du nombre des sinistres si on s'intéresse au nombre moyen de sinistres, ou par la suite du coût des sinistres si on s'intéresse au coût unitaire moyen des sinistres.

Pour une catégorie $x_i \in \mathbb{R}^d$ fixée, afin d'alléger les notations, on introduit l'opérateur $\mathbb{E}_{x_i} := \mathbb{E}[\cdot \mid \{X_i = x_i\}]$, c'est à dire l'espérance conditionnellement à $\{X_i = x_i\}$. On pose :

$$P(\theta_i) := \mathbb{E}_{x_i} [S_{t+1}^i - S_t^i \mid \Theta_i = \theta_i],$$

$$\bar{S}_t^i := \frac{S_t^i}{t}.$$

Par construction, $\hat{P}_i = \mathbb{E}_{x_i} [P(\Theta_i)]$. La crédibilité non paramétrique consiste à résoudre le problème :

$$\min_{(a,b) \in \mathbb{R}^2} \mathbb{E}_{x_i} \left[\left(P(\Theta_i) - a - b\bar{S}_t^i \right)^2 \right].$$

Les conditions du premier ordre donnent :

$$\begin{cases} -2\mathbb{E}_{x_i} \left[P(\Theta_i) - a^* - b^*\bar{S}_t^i \right] = 0 \\ -2\mathbb{E}_{x_i} \left[\bar{S}_t^i \left(P(\Theta_i) - a^* - b^*\bar{S}_t^i \right) \right] = 0 \end{cases} \iff \begin{cases} \mathbb{E}_{x_i} \left[P(\Theta_i) - a^* - b^*\bar{S}_t^i \right] = 0 \\ \mathbb{E}_{x_i} \left[\bar{S}_t^i \left(P(\Theta_i) - a^* - b^*\bar{S}_t^i \right) \right] = 0 \end{cases}$$

La première équation donne : $a^* = \mathbb{E}_{x_i} [P(\Theta_i) - b^*\bar{S}_t^i]$. En substituant dans a^* la seconde, on obtient :

$$\mathbb{E}_{x_i} \left[\bar{S}_t^i P(\Theta_i) \right] - \mathbb{E}_{x_i} \left[P(\Theta_i) - b^*\bar{S}_t^i \right] \mathbb{E}_{x_i} \left[\bar{S}_t^i \right] - b^* \mathbb{E}_{x_i} \left[\left(\bar{S}_t^i \right)^2 \right] = 0$$

$$\iff Cov_{x_i} \left[\bar{S}_t^i, P(\Theta_i) \right] - b^* Var_{x_i} (\bar{S}_t^i) = 0.$$

On retrouve les résultats de la régression linéaire simple, c'est-à-dire :

$$a^* = \mathbb{E}_{x_i} [P(\Theta_i)] - b^* \mathbb{E}_{x_i} [\bar{S}_t^i],$$

$$b^* = \frac{\text{Cov}_{x_i} [\bar{S}_t^i, P(\Theta_i)]}{\text{Var}_{x_i} (\bar{S}_t^i)}.$$

On pose $Z_t := S_t^i - S_{t-1}^i$ pour $t \in \mathbb{N}^*$.

$$\begin{aligned} \text{Cov}_{x_i} [\bar{S}_t^i, P(\Theta_i)] &= \frac{1}{t} \sum_{j=1}^t \text{Cov}_{x_i} [Z_j, P(\Theta_i)] = \text{Cov}_{x_i} [Z_t, P(\Theta_i)] \\ &= \mathbb{E}_{x_i} [Z_t P(\Theta_i)] - \mathbb{E}_{x_i} [Z_t] \mathbb{E}_{x_i} [P(\Theta_i)] \\ &= \mathbb{E}_{x_i} [\mathbb{E}_{x_i} [Z_t P(\Theta_i) \mid \Theta_i]] - \mathbb{E}_{x_i} [\mathbb{E}_{x_i} [Z_t \mid \Theta_i]] \mathbb{E}_{x_i} [P(\Theta_i)] \\ &= \mathbb{E}_{x_i} [P(\Theta_i) \mathbb{E}_{x_i} [Z_t \mid \Theta_i]] - \mathbb{E}_{x_i} [P(\Theta_i)]^2 \\ &= \text{Var}_{x_i} [P(\Theta_i)]. \end{aligned}$$

On obtient

$$\begin{aligned} P &= a^* + b^* \bar{S}_t^i = \mathbb{E}_{x_i} [P(\Theta_i)] + \frac{\text{Cov}_{x_i} [\bar{S}_t^i, P(\Theta_i)]}{\text{Var}_{x_i} (\bar{S}_t^i)} \left(\bar{S}_t^i - \mathbb{E}_{x_i} [\bar{S}_t^i] \right) \\ &= \mathbb{E}_{x_i} [P(\Theta_i)] + \frac{\text{Var}_{x_i} [P(\Theta_i)]}{\text{Var}_{x_i} (\bar{S}_t^i)} \left(\bar{S}_t^i - \mathbb{E}_{x_i} [P(\Theta_i)] \right) \\ &= \left(1 - \frac{\text{Var}_{x_i} [P(\Theta_i)]}{\text{Var}_{x_i} (\bar{S}_t^i)} \right) \mathbb{E}_{x_i} [P(\Theta_i)] + \frac{\text{Var}_{x_i} [P(\Theta_i)]}{\text{Var}_{x_i} (\bar{S}_t^i)} \bar{S}_t^i. \end{aligned}$$

On a :

$$\begin{aligned} \text{Var}_{x_i} (\bar{S}_t^i) &= \mathbb{E}_{x_i} [\text{Var}_{x_i} (\bar{S}_t^i \mid \Theta_i)] + \text{Var}_{x_i} [\mathbb{E}_{x_i} (\bar{S}_t^i \mid \Theta_i)] \\ &= \frac{\text{Var}_{x_i} (Z_t \mid \Theta_i)}{t} + \text{Var}_{x_i} [P(\Theta_i)]. \end{aligned}$$

On pose :

$$z_t := \frac{t}{t + \frac{\text{Var}_{x_i}(Z_t \mid \Theta_i)}{\text{Var}_{x_i}[P(\Theta_i)]}}.$$

Alors :

$$P = (1 - z_t) \mathbb{E}_{x_i} [P(\Theta_i)] + z_t \bar{S}_t^i.$$

Et en réintroduisant la caractéristique $X_i = x_i$,

$$P = (1 - z_t) \mathbb{E} [P(\Theta_i) \mid X = x_i] + z_t \bar{S}_t^i = (1 - z_t) \mathbb{E} [P(\Theta_i) \mid X = x_i] + z_t \bar{S}_t^i.$$

Dans le cas particulier du nombre de sinistres, cela s'écrit :

$$P_t^N = (1 - z_t) \hat{P}_i^N[x_i] + z_t \bar{N}_t^i.$$

Deuxième partie

Sinitres tardifs

En assurance, les contrats couvrent les sinistres sur une année. L'assureur encaisse la prime immédiatement et est amené à payer par après les éventuels sinistres. Toutefois, à la fin des contrats, l'assureur ne sait pas exactement ce qu'il devra payer en conséquence de deux cas :

- Certains sinistres n'ont pas un coût final complètement déterminé,
- Certains sinistres ont eu lieu durant la période de couverture mais ne sont pas encore connus.

Ces deux cas amènent l'assureur à devoir estimer le coût réel de ses garanties passées échues car il ne peut pas connaître avec exactitude leur valeur. Ce sont les sinistres tardifs qu'on appelle en anglais les IBNR (Incurred But Not Reported), sachant que dans le premier cas le sinistre est connu, seul le montant est incertain.

Pour les contrats signés il y a $n \in \mathbb{N}^*$ années, l'assureur a pu observer le flux d'information pendant n années, ce qui lui permet :

- D'avoir une bonne connaissance des sinistres tardifs,
- De connaître plus précisément le coût des années précédentes.

3 Les triangles de liquidation

On se place sur un espace probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On définit les variables aléatoire réelles $X_{i,j}$ pour $1 \leq i, j \leq n$ où $n \in \mathbb{N}^*$ est un nombre d'années.

- i correspond à l'année de survenance du sinistre, il s'agit de l'année de couverture du contrat
- j correspond à l'année de développement du sinistres, il s'agit de l'année à laquelle l'information arrive à l'assureur, par rapport à i (par exemple, $j = 1$ correspond à l'année de couverture du contrat, $j = 2$ à un an après la couverture du contrat)
- $X_{i,j}$ est la variable aléatoire qui correspond au coût des sinistres survenus l'année i pour l'année de développement j .

La v.a. $X_{i,j}$ représente :

- Si $j = 1$, le coût estimé des sinistres survenus en l'année i et connus à la fin de l'année,
- Si $j > 1$, la variation du coût estimé des sinistres survenus en l'année i et appris à la j -ème année

Par construction, en une année $k \in \mathbb{N}^*$, on observe les variables aléatoires $X_{i,j}$ telles que $i + j \leq k + 1$. On introduit la filtration :

$$\mathcal{F}_k := \sigma(X_{i,j} \mid i + j \leq k + 1).$$

Aujourd'hui, en année n , les variables observées sont celles qui sont \mathcal{F}_n -mesurables, c'est à dire les $X_{i,j}$ tels que $i + j \leq n + 1$ qui sont représentées par le triangle :

i, j	1	\dots	n
1	$X_{1,1}$	\dots	$X_{1,n}$
\vdots	\vdots	\ddots	
n	$X_{n,1}$		

On définit, pour $1 \leq i, j \leq n$,

$$C_{i,j} = \sum_{k=1}^j X_{i,k}.$$

Il s'agit de la somme cumulée et par construction, $C_{i,j}$ est \mathcal{F}_k -mesurable si $i + j \leq k + 1$, comme $X_{i,j}$. Il s'agit du coût total des sinistres survenus l'année i et vus en l'année de développement j . Si on suppose que les sinistres se développent en maximum n années, alors $C_{i,n}$ est le coût total des sinistres survenus pendant l'année i .

Nous avons le triangle cumulé :

i, j	1	\dots	n
1	$C_{1,1}$	\dots	$C_{1,n}$
\vdots	\vdots	\ddots	
n	$C_{n,1}$		

L'objectif principal est alors d'estimer la partie inconnue du triangle et d'en déduire une approximation du montant restant à payer pour les contrats passés.

3.1 L'exposition

L'exposition correspond à une mesure du risque du portefeuille.

On note $E_i \in \mathbb{R}_+$ l'exposition de l'année i . Plus l'exposition sera élevée, plus l'assureur aura souscrit de contrats et sera donc exposé à des sinistres tardifs.

Si tous les contrats sont identiques, l'exposition est le nombre de contrats en cours l'année i . Le montant total des primes est une mesure de l'exposition qui tient compte de la différence des risques des différents contrats.

Certains contrats sont signés en cours d'année, d'autres s'arrêtent en cours d'année. Il est alors possible de définir une exposition individuel $(e_{k,i})_{k \geq 1}$ d'un contrat pour une année i et un individu k . Si le contrat a couvert une proportion $p_{k,i} \in [0, 1]$ de l'année i , l'exposition associée au contrat (dans le cas où on compte le nombre de contrats) est :

$$e_{i,k} := p_{i,k}.$$

Si m est le nombre total de contrats, l'exposition totale de l'année i est :

$$E_i := \sum_{k=1}^m e_{i,k}.$$

L'exposition est une notion importante. En effet, si l'assureur a un nombre de contrats supérieur l'année $i + 1$ par rapport à l'année i , on s'attend observer en moyenne des valeurs plus élevées pour $(X_{i,j})_{j \geq 1}$.

3.2 Risque de réserve

Le premier objectif est d'être capable d'estimer la partie inconnue du triangle en l'année n , c'est à dire les $X_{i,j}$ pour $i + j > n$. Cela permet d'estimer le coût total $(C_{i,n})_{1 \leq i \leq n}$ de chaque année. Le montant à prévoir pour les évolution futur est $C_{i,n} - C_{i,n+1-i}$. La réserve est alors :

$$R := \sum_{i=1}^n (C_{i,n} - C_{i,n+1-i})$$

C'est une variable aléatoire pour laquelle nous mettrons en place des techniques afin d'estimer la moyenne compte tenu de l'information disponible, ainsi que le risque associé (car même si nous étions capable d'en déduire la moyenne exacte vu de n , il ne faut pas oublier qu'il s'agit d'une variable aléatoire).

3.3 Risque de souscription

Le risque de souscription pour la nouvelle année $n+1$ est la variable aléatoire $C_{n+1,n}$. Sachant l'exposition E_{n+1} , on peut estimer le risque associé à la souscription pour l'année $n + 1$ en s'intéressant à la distribution de $C_{n+1,n}$.

4 Modèles de Chain Ladder - Mack

Cette section s'appuie principalement sur [2].

4.1 Estimation des réserves

Il s'agit du modèle le plus utilisé en assurance. Il repose sur les hypothèses suivantes :

Hypothèse 2.

H1 Les v.a. $(C_{i_1,j})_{1 \leq j \leq n}$ et $(C_{i_2,j})_{1 \leq j \leq n}$ sont indépendantes pour $i_1 \neq i_2$.

H2 Pour $1 \leq j \leq n - 1$, il existe $f_j \geq 0$ tel que

$$\mathbb{E}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) = f_j C_{i,j}.$$

H3 Pour $1 \leq j \leq n - 1$, il existe $\sigma_j \geq 0$ tel que

$$\text{Var}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) = \sigma_j^2 C_{i,j}.$$

Ainsi, on fait trois hypothèses. La première, c'est que les années de survenance des sinistres sont indépendantes. La deuxième est que la variation du montant à payer en moyenne pour l'année i lors du développement $j + 1$ est proportionnelle à l'année j . Enfin, la variance est proportionnelle à $C_{i,j}$ ce qui implicitement revient à avoir un grand portefeuille diversifié.

On introduit la filtration $\mathcal{B}_k = \sigma(C_{i,j} \mid j \leq k)$.

Remarque 1. *Sous H1, H2 et H3 vérifient également :*

$$\begin{aligned}\mathbb{E}(C_{i,j+1} \mid \mathcal{B}_j) &= f_j C_{i,j}, \\ \text{Var}(C_{i,j+1} \mid \mathcal{B}_j) &= \sigma_j^2 C_{i,j}.\end{aligned}$$

Démonstration. Par H1, à i fixé, les seules v.a. qui interviennent dans le conditionnement par \mathcal{F}_{i+j-1} sont $(C_{i,k})_{1 \leq k \leq j}$. Par cette même hypothèse, ce sont les seules variables qui interviennent également dans le conditionnement par \mathcal{B}_j ce qui permet d'en déduire le résultat. \square

On peut poser $\hat{f}_{i,j} := \frac{C_{i,j+1}}{C_{i,j}}$ pour $i + j + 1 \leq n + 1$, i.e. pour les cas observables en année n . Un estimateur de f_j est :

$$\hat{f}_j := \frac{1}{n-j} \sum_{i=1}^{n-j} \hat{f}_{i,j}.$$

Toutefois, il ne s'agit pas du meilleur estimateur linéaire de f_j avec les observations $\hat{f}_{i,j}$ au sens de l'erreur quadratique moyenne.

Avec les poids positifs $p := (p_{i,j})_{1 \leq i \leq n, 1 \leq j \leq n-i}$ (et non tous nuls), on redéfinit :

$$\hat{f}_j := \frac{\sum_{i=1}^{n-j} p_{i,j} \hat{f}_{i,j}}{\sum_{i=1}^{n-j} p_{i,j}}.$$

Lemme 4. *Pour tout $1 \leq j \leq n-1$, on a :*

$$\mathbb{E}(\hat{f}_{i,j} \mid \mathcal{B}_j) = f_j.$$

Démonstration. Par H2 :

$$\mathbb{E}(\hat{f}_{i,j} \mid \mathcal{B}_j) = \frac{1}{C_{i,j}} \mathbb{E}(C_{i,j+1} \mid \mathcal{B}_j) = f_j.$$

\square

Proposition 1. *Sous H2, pour tout $1 \leq j \leq n-1$, pour tous les poids p qui sont \mathcal{B}_j -mesurable, l'estimateur \hat{f}_j est un estimateur sans biais de f_j conditionnellement à l'information disponible \mathcal{B}_j , c'est-à-dire :*

$$\mathbb{E}(\hat{f}_j \mid \mathcal{B}_j) = f_j.$$

En particulier, il est sans biais :

$$\mathbb{E}(\hat{f}_j) = f_j.$$

Démonstration.

$$\mathbb{E}(\hat{f}_j \mid \mathcal{B}_j) = \mathbb{E}\left(\frac{\sum_{i=1}^{n-j} p_{i,j} \hat{f}_{i,j}}{\sum_{i=1}^{n-j} p_{i,j}} \mid \mathcal{B}_j\right) = \frac{\sum_{i=1}^{n-j} p_{i,j} \mathbb{E}(\hat{f}_{i,j} \mid \mathcal{B}_j)}{\sum_{i=1}^{n-j} p_{i,j}} = f_j.$$

\square

Il reste à choisir les poids optimaux. On va s'appuyer sur le résultat suivant.

Lemme 5. Soit \mathcal{B} une tribu. Soient (X_1, \dots, X_n) n variables aléatoires indépendantes de L^2 de variance conditionnelle $\text{Var}(X_i | \mathcal{B}) = \sigma_i^2$. Soit :

$$\hat{\theta}_n := \sum_{i=1}^n p_i X_i,$$

avec $\sum_{i=1}^n p_i = 1$ qui sont \mathcal{B} -mesurables. Le choix des $(p_i)_{1 \leq i \leq n}$ qui minimise $\text{Var}(\hat{\theta}_n | \mathcal{B})$ est :

$$p_i \propto \frac{1}{\sigma_i^2}.$$

Démonstration. On souhaite minimiser sur p :

$$\text{Var}(\hat{\theta}_n | \mathcal{B}) + \lambda \left(\sum_{i=1}^n p_i - 1 \right) = \sum_{i=1}^n p_i^2 \sigma_i^2 + \lambda \left(\sum_{i=1}^n p_i - 1 \right).$$

La condition du premier ordre donne :

$$2p_i \sigma_i^2 + \lambda = 0 \iff p_i = -\frac{\lambda}{2\sigma_i^2},$$

c'est-à-dire que :

$$p_i \propto \frac{1}{\sigma_i^2}.$$

□

Proposition 2. Sous $H1 - H2 - H3$, le meilleur estimateur linéaire de f_j en $(\hat{f}_{i,j})_{1 \leq i \leq n-j}$ pour l'erreur quadratique moyenne conditionnellement à \mathcal{B}_j est

$$\hat{f}_j := \frac{\sum_{i=1}^{n-j} C_{i,j} \hat{f}_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}},$$

c'est-à-dire :

$$\hat{f}_j := \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}.$$

Démonstration. Pour $1 \leq j \leq n-1$, l'erreur quadratique moyenne est :

$$\mathbb{E} \left[(\hat{f}_j - f_j)^2 | \mathcal{B}_j \right] = \left[\mathbb{E}(\hat{f}_j | \mathcal{B}_j) - f_j \right]^2 + \text{Var} \left[\hat{f}_j | \mathcal{B}_j \right]$$

Comme l'estimateur est sans biais, cela revient à minimiser la variance. Par H3 :

$$\text{Var}(\hat{f}_{i,j} | \mathcal{B}_j) = \frac{\sigma_j^2}{C_{i,j}}.$$

Par le Lemme 5, on en déduit que :

$$p_{i,j} \propto C_{i,j}.$$

□

On peut en déduire la variance des \hat{f}_j .

Lemme 6. On a

$$\text{Var}(\hat{f}_j \mid \mathcal{B}_j) = \frac{\sigma_j^2}{\sum_{i=1}^{n-j} C_{i,j}}$$

Démonstration. On a par H1,

$$\text{Var}(\hat{f}_j \mid \mathcal{B}_j) = \frac{\sum_{i=1}^{n-j} C_{i,j}^2 \text{Var}(\hat{f}_{i,j} \mid \mathcal{B}_j)}{\left(\sum_{i=1}^{n-j} C_{i,j}\right)^2} = \frac{\sum_{i=1}^{n-j} C_{i,j}^2 \frac{\sigma_j^2}{C_{i,j}}}{\left(\sum_{i=1}^{n-j} C_{i,j}\right)^2} = \sigma_j^2 \frac{\sum_{i=1}^{n-j} C_{i,j}}{\left(\sum_{i=1}^{n-j} C_{i,j}\right)^2} = \frac{\sigma_j^2}{\sum_{i=1}^{n-j} C_{i,j}}$$

□

Un autre résultat est que les estimateurs \hat{f}_j ne sont pas corrélés entre eux.

Lemme 7. Pour $j \neq j'$, on a

$$\mathbb{E}(\hat{f}_j \hat{f}_{j'}) = \mathbb{E}(\hat{f}_j) \mathbb{E}(\hat{f}_{j'}).$$

Démonstration. Si $j < j'$

$$\mathbb{E}[\mathbb{E}(\hat{f}_j \hat{f}_{j'} \mid \mathcal{B}_{j'})] = \mathbb{E}[\hat{f}_j \mathbb{E}(\hat{f}_{j'} \mid \mathcal{B}_{j'})] = \mathbb{E}[\hat{f}_j] f_{j'} = \mathbb{E}[\hat{f}_j] \mathbb{E}[\hat{f}_{j'}].$$

□

On a le lemme suivant :

Lemme 8. On pose $j = n - i + 1$ (qui est la dernière observation de $C_{i,j}$ pour l'année i).

$$\mathbb{E}(C_{i,n} \mid \mathcal{F}_n) = \left(\prod_{k=j}^{n-1} f_k \right) C_{i,j}.$$

Démonstration. Pour $j < n$, on a

$$\mathbb{E}(C_{i,n} \mid \mathcal{F}_n) = \mathbb{E}(C_{i,n} \mid \mathcal{B}_j) = \mathbb{E}(\mathbb{E}(C_{i,n} \mid \mathcal{B}_{n-1}) \mid \mathcal{B}_j) = f_{n-1} \mathbb{E}(C_{i,n-1} \mid \mathcal{F}_n).$$

Puis, par récurrence, on en déduit le résultat. □

Proposition 3. On pose $j = n - i + 1$ (qui est la dernière observation de $C_{i,j}$ pour l'année i). L'estimateur

$$\hat{C}_{i,n} := \left(\prod_{k=j}^{n-1} \hat{f}_k \right) C_{i,j}$$

est un estimateur sans biais de $C_{i,n}$.

Démonstration. Admis. □

Le résultat précédent permet d'en déduire un estimateur sans biais des réserves associées à chaque année de sinistralité qui est :

$$\hat{R}_i := \hat{C}_{i,n} - C_{i,n-i+1}.$$

Et enfin, un estimateur du besoin en réserve total :

$$\hat{R} := \sum_{i=2}^n \hat{R}_i.$$

Remarque 2. Le modèle est invariant par changement d'unité ou de monnaie. En effet, si on multiplie tous les triangles par $\alpha > 0$, on remarque qu'on a les mêmes estimateurs et la même estimation des réserves.

4.2 Mesure de l'erreur sur l'estimation des réserves

On sait que les \hat{f}_j , les $\hat{C}_{i,n}$ et les \hat{R}_i sont des estimateurs centrés, mais demeurent des variables aléatoires. Nous allons donc calculer une mesure de l'erreur : l'erreur quadratique moyenne entre \hat{R}_i et R_i dans un premier temps puis entre \hat{R} et R dans un second temps.

Nous aurons besoin d'un estimateur des σ_j^2 .

Lemme 9. *Supposons que nous connaissons les $(f_j)_{1 \leq j \leq n-1}$. Pour $1 \leq j \leq n-1$, l'estimateur :*

$$\hat{\sigma}_j^2 := \frac{1}{n-j} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - f_j \right)^2 = \frac{1}{n-j} \sum_{i=1}^{n-j} C_{i,j} \left(\hat{f}_{i,j} - f_j \right)^2$$

est un estimateur sans biais de σ_j^2 .

Démonstration. On rappelle H3, pour $1 \leq j \leq n-1$:

$$\text{Var}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) = \sigma_j^2 C_{i,j}.$$

C'est à dire que :

$$\begin{aligned} & \mathbb{E} \left[(C_{i,j+1} - \mathbb{E}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}))^2 \mid \mathcal{F}_{i+j-1} \right] = \sigma_j^2 C_{i,j} \\ \iff & \mathbb{E} \left[(C_{i,j+1} - f_j C_{i,j})^2 \mid \mathcal{F}_{i+j-1} \right] = \sigma_j^2 C_{i,j} \\ \iff & \mathbb{E} \left[C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - f_j \right)^2 \mid \mathcal{F}_{i+j-1} \right] = \sigma_j^2. \end{aligned}$$

□

Comme nous n'avons aucune hypothèse supplémentaire, l'estimateur est la moyenne pondérée associée. Mais nous ne connaissons pas les f_j , nous devons les remplacer par leurs estimateurs respectifs.

Lemme 10. *Pour $1 \leq j \leq n-2$, l'estimateur :*

$$\hat{\sigma}_j^2 := \frac{1}{n-j-1} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2$$

est un estimateur sans biais de σ_j^2 .

Démonstration. Admis. □

On remarque qu'on ne peut pas avoir d'estimateur de σ_{n-1}^2 . On pourra utiliser une extrapolation, on suppose pour la suite avoir à disposition un tel estimateur. Par exemple, [2] propose :

$$\hat{\sigma}_{n-1}^2 := \min(\hat{\sigma}_{n-2}^4 / \hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2).$$

Proposition 4. *L'erreur quadratique moyenne des réserves pour une année i estimée \hat{R}_i , notée $\mathbb{E} \left[(\hat{R}_i - R_i)^2 \mid \mathcal{F}_n \right]$, peut être estimée par :*

$$\widehat{\text{eqm}}(\hat{R}_i) = \hat{C}_{i,n}^2 \sum_{j=n+1-i}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{n-j} C_{k,j}} \right).$$

Démonstration. Admis. □

La première partie de l'erreur quadratique moyenne correspond au fait que, vu d'aujourd'hui, $C_{i,n}$ est une variable aléatoire. Si nous connaissions les vrais paramètres, elle serait inconnue d'aujourd'hui et cela correspond à sa variance. La seconde partie correspond à une erreur d'estimation des paramètres.

Proposition 5. *L'erreur quadratique moyenne des réserves totales peut être estimée par :*

$$\widehat{eqm(\hat{R})} = \sum_{i=2}^n \left(eqm(\hat{R}_i) + \hat{C}_{i,n} \left(\sum_{k=i+1}^n \hat{C}_{k,n} \right) \sum_{j=n+1-i}^{n-1} \frac{2\hat{\sigma}_j^2 / \hat{f}_j^2}{\sum_{\ell=1}^{n-j} C_{\ell,j}} \right).$$

Démonstration. Admis. □

La première partie (somme des erreurs quadratiques moyennes) correspond à la somme des différentes années, indépendantes. La seconde est associée à l'erreur des paramètres : ce sont les mêmes paramètres qui interviennent, la diversification est amoindrie.

4.3 Exemple

Nous avons le triangle cumulé suivant (tiré de [2]).

i	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$	$C_{i,4}$	$C_{i,5}$	$C_{i,6}$	$C_{i,7}$	$C_{i,8}$	$C_{i,9}$	$C_{i,10}$
1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
5	443160	1136350	2128333	2897821	3402672	3873311				
6	396132	1333217	2180715	2985752	3691712					
7	440832	1288463	2419861	3483130						
8	359480	1421128	2864494							
9	376686	1363294								
10	344014									

Si le triangle ci-dessus est importé dans une matrice ou data.frame qui s'appelle C dans le langage R, alors le code suivant permet de calculer $(\hat{f}_j)_{1 \leq j \leq n-1}$.

```
f <- numeric(n-1L)
for(j in 1L:(n-1L))
  f[j] <- sum(C[1L:(n-j), j + 1L]) / sum(C[1L:(n-j), j])
```

Le code suivant permet de calculer $(\hat{\sigma}_j^2)_{1 \leq j \leq n-2}$.

```
s2 <- numeric(n-1L)
for(j in 1L:(n-2L))
  s2[j] <- sum( C[1L:(n-j), j] * (C[1L:(n-j), j + 1L] / C[1L:(n-j), j]
    - f[j])^2 ) / (n - j - 1L)
```

Enfin, on utilise l'extrapolation de [2] pour estimer σ_{n-1}^2 .

```
s2[n-1L] <- min(s2[n-2L]^2 / s2[n-3L], s2[n-3L], s2[n-2L])
```

On obtient le Tableau 1.

On peut ensuite estimer les $C_{i,j}$ pour $i + j > n + 1$ avec le code R suivant.

j	1	2	3	4	5	6	7	8	9
\hat{f}_j	3.491	1.747	1.457	1.174	1.104	1.086	1.054	1.077	1.018
$\hat{\sigma}_j^2$	160280	37737	41965	15183	13731	8186	447	1147	447

TABLE 1 – Estimateurs des paramètres de Mack - Chain Ladder

```
for(i in 2L:n)
  C[i, (n-i+2L):n] <- cumprod(f[(n-i+1L):(n-1L)])*C[i, n - i + 1L]
```

Et en déduire les réserves.

```
R <- C[, n] - rev(C[row(C) + col(C) == n + 1L])
```

On obtient :

i	$\hat{C}_{i,n}$	\hat{R}_i
1	3901463	0
2	5433719	94634
3	5378826	469511
4	5297906	709638
5	4858200	984889
6	5111171	1419459
7	5660771	2177641
8	6784790	3920296
9	5642265	4278971
10	4969824	4625810

La réserve totale est $\hat{R} = 18680848$. Nous pouvons maintenant calculer l'erreur quadratique moyenne de chaque année avec le code suivant.

```
C0 <- C; C0[row(C) + col(C) >= n+1L] <- 0
eqm_R <- numeric(n)
for(i in 2L:n)
{
  J <- (n+1L-i):(n-1L)
  eqm_R[i] <- C[i, n]^2*sum( (s2[J]/f[J]^2) *
    (1/C[i, J] + 1/colSums(C0[, J, drop = FALSE])) )
}
```

En exprimant sa racine carrée, pour chaque année, en pourcentage de l'estimateur des réserves, on obtient :

i	$\frac{\sqrt{EQM(\hat{R}_i)}}{\hat{R}_i} (\%)$
1	0.0
2	79.8
3	25.9
4	18.8
5	26.5
6	29.0
7	25.6
8	22.3
9	22.7
10	29.5

Enfin, le code R suivant permet de calculer l'erreur quadratique moyenne d'estimation des réserves :

```
eqm <- eqm_R[n]
for(i in 2L:(n-1L))
{
  J <- (n+1L-i):(n-1L)
  eqm <- eqm + eqm_R[i] + C[i, n]*sum(C[(i+1):n, n]) *
    sum((2*s2[J]/f[J]^2)/colSums(C0[, J, drop = FALSE]))
}
```

En prenant la racine carrée de l'erreur quadratique moyenne, nous obtenons 13.1% par rapport au montant estimé des réserves.

5 Modèle de Schnieper

Cette section s'appuie principalement sur [3].

5.1 Estimation des réserves

Cette fois-ci nous allons utiliser l'exposition. L'idée est la suivante : le défaut de Chain Ladder est le caractère multiplicatif pur : les sinistres latents sont proportionnels à ce qui est connu. L'idée du modèle de Schnieper est la suivante :

- Il y a les sinistres qui ont déjà eu lieu et qui sont inconnus de l'assureur, ils sont proportionnels à l'exposition, il s'agit des *vrais IBNR* ;
- Il y a les sinistres connus mais dont le montant estimé peut évoluer, cette évolution est proportionnelle à la charge actuelle, il s'agit de révision de charge de sinistre : les *IBNER (Incurred But Not Enough Reported)*.

On reprend les variables aléatoires $(C_{i,j})_{1 \leq i, j \leq n}$ qui représentent le coût total cumulé. Nous définissons de plus :

- Les variables aléatoires $(N_{i,j})_{1 \leq i, j \leq n}$ qui représentent le montant des nouveaux sinistres survenus en année i et déclarés en année de développement j ;
- Les variables aléatoires $(D_{i,j})_{1 \leq i, j \leq n}$ qui représentent la baisse du coût estimé des sinistres déclarés les années précédentes.

Par construction $D_{i,1} = 0$ pour tout $1 \leq i \leq n$ et $D_{i,j} \leq C_{i,j-1}$ pour $1 \leq i \leq n$ et $2 \leq j \leq n$. Le coût incrémental des sinistres s'écrit :

$$\begin{aligned} C_{i,1} &= N_{i,1}, & 1 \leq i \leq n \\ C_{i,j+1} &= C_{i,j} + N_{i,j+1} - D_{i,j+1}, & 1 \leq i \leq n, 1 \leq j \leq n-1 \end{aligned} \tag{1}$$

Parmi C, N, D , il suffit d'en connaître deux pour en déduire le troisième, l'un d'entre-eux est redondant. On introduit la filtration :

$$\mathcal{F}_k := \sigma(N_{i,j}, D_{i,j} \mid i + j \leq k + 1).$$

Nous avons donc deux triangles de données, dont l'agrégation en suivant (1) donne C .

Le modèle repose sur les hypothèses suivantes :

Hypothèse 3.

H1 Les v.a. $(N_{i_1,j}, D_{i_1,j})_{1 \leq j \leq n}$ et $(N_{i_2,j}, D_{i_2,j})_{1 \leq j \leq n}$ sont indépendantes pour $i_1 \neq i_2$.

H2 Pour $1 \leq j \leq n$, il existe $\lambda_j \geq 0$ et pour $1 \leq j \leq n-1$ il existe $\delta_j \leq 1$ tel que

$$\begin{aligned}\mathbb{E}(N_{i,j} \mid \mathcal{F}_{i+j-2}) &= \lambda_j E_i, \\ \mathbb{E}(D_{i,j+1} \mid \mathcal{F}_{i+j-1}) &= \delta_j C_{i,j}.\end{aligned}$$

H3 Pour $1 \leq j \leq n-1$, il existe $\sigma_j \geq 0$ et $\tau_j^2 \geq 0$ tel que

$$\begin{aligned}\text{Var}(N_{i,j} \mid \mathcal{F}_{i+j-2}) &= \sigma_j^2 E_i, \\ \text{Var}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) &= \tau_j^2 C_{i,j}.\end{aligned}$$

Remarque 3. Par la formule des espérances conditionnelles emboîtées et de la variance totale, en combinant H2 et H3, on a également :

$$\begin{aligned}\mathbb{E}(N_{i,j}) &= \lambda_j E_i, \\ \text{Var}(N_{i,j}) &= \sigma_j^2 E_i.\end{aligned}$$

On introduit la filtration $\mathcal{B}_k = \sigma(C_{i,j} \mid j \leq k)$.

Remarque 4. Sous H1, H2 et H3 vérifient également :

$$\begin{aligned}\mathbb{E}(N_{i,j} \mid \mathcal{B}_{j-1}) &= \lambda_j E_i, \\ \mathbb{E}(D_{i,j+1} \mid \mathcal{B}_j) &= \delta_j C_{i,j}, \\ \text{Var}(D_{i,j+1} \mid \mathcal{B}_j) &= \tau_j^2 C_{i,j}.\end{aligned}$$

Démonstration. Voir la preuve de la Remarque 1. □

On peut poser $\hat{\lambda}_{i,j} := \frac{N_{i,j}}{E_i}$ pour $i+j \leq n+1$, i.e. pour les cas observables en année n ; et $\hat{\delta}_{i,j} := \frac{D_{i,j+1}}{C_{i,j}}$ pour $i+j+1 \leq n+1$.

En introduisant des poids comme dans le modèle Chain Ladder, un estimateur de λ_j et un estimateur de δ_j sont :

$$\begin{aligned}\hat{\lambda}_j &:= \frac{\sum_{i=1}^{n-j+1} p_{i,j} \hat{\lambda}_{i,j}}{\sum_{i=1}^{n-j} p_{i,j}}, & 1 \leq j \leq n, \\ \hat{\delta}_j &:= \frac{\sum_{i=1}^{n-j} p_{i,j} \hat{\delta}_{i,j}}{\sum_{i=1}^{n-j} p_{i,j}}, & 1 \leq j \leq n-1.\end{aligned}$$

Lemme 11. Pour tout $1 \leq j \leq n-1$, on a :

$$\begin{aligned}\mathbb{E}(\hat{\lambda}_{i,j} \mid \mathcal{B}_{j-1}) &= \lambda_j, \\ \mathbb{E}(\hat{\delta}_{i,j} \mid \mathcal{B}_j) &= \delta_j.\end{aligned}$$

Démonstration. □

Proposition 6. Sous H1 – H2 – H3, le meilleur estimateur linéaire de λ_j en $(\hat{\lambda}_{i,j})_{1 \leq i \leq n-j+1}$ et le meilleur estimateur linéaire de δ_j en $(\hat{\delta}_{i,j})_{1 \leq i \leq n-j}$ pour l'erreur quadratique moyenne conditionnellement à \mathcal{B}_j sont

$$\begin{aligned}\hat{\lambda}_j &:= \frac{\sum_{i=1}^{n-j+1} E_i \hat{\lambda}_{i,j}}{\sum_{i=1}^{n-j+1} E_i}, \\ \hat{\delta}_j &:= \frac{\sum_{i=1}^{n-j} C_{i,j} \hat{\delta}_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}},\end{aligned}$$

c'est-à-dire :

$$\hat{\lambda}_j := \frac{\sum_{i=1}^{n-j+1} N_{i,j}}{\sum_{i=1}^{n-j+1} E_i},$$

$$\hat{\delta}_j := \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}.$$

Démonstration. □

On peut en déduire la variance des $\hat{\lambda}_j$ et des $\hat{\delta}_j$.

Lemme 12. On a

$$\text{Var}(\hat{\lambda}_j | \mathcal{B}_j) = \frac{\sigma_j^2}{\sum_{i=1}^{n-j+1} E_i},$$

$$\text{Var}(\hat{\delta}_j | \mathcal{B}_j) = \frac{\tau_j^2}{\sum_{i=1}^{n-j} C_{i,j}}.$$

Démonstration. □

Lemme 13. On pose $j = n - i + 1$ (qui est la dernière observation de $C_{i,j}$ pour l'année i).

$$\mathbb{E}(C_{i,n} | \mathcal{F}_n) = \left(\prod_{k=j}^{n-1} (1 - \delta_k) \right) C_{i,j} + \sum_{k=j+1}^n \lambda_k E_i \left(\prod_{\ell=k}^{n-1} (1 - \delta_\ell) \right).$$

Démonstration. □

Remarque 5. Le modèle est invariant par changement d'unité ou de monnaie. En effet, si on multiplie tous les triangles ainsi que les expositions $(E_i)_{1 \leq i \leq n}$ par $\alpha > 0$, on remarque qu'on a les mêmes estimateurs et la même estimation des réserves. Si on ne multiplie par E_i par α (c'est, par exemple, le nombre de contrat), nous n'avons plus les mêmes estimateurs de $\hat{\lambda}_j$ mais l'estimation des réserves est inchangée.

5.2 Mesure de l'erreur

Lemme 14. Supposons que nous connaissons les $(\lambda_k)_{1 \leq k \leq n}$ et les $(\delta_k)_{1 \leq k \leq n-1}$. Pour $1 \leq j \leq n - 1$, les estimateurs :

$$\hat{\sigma}_j^2 := \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} E_i \left(\frac{N_{i,j}}{E_i} - \lambda_j \right)^2 = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} E_i \left(\hat{\lambda}_{i,j} - \lambda_j \right)^2,$$

$$\hat{\tau}_j^2 := \frac{1}{n-j} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{D_{i,j+1}}{C_{i,j}} - \delta_j \right)^2 = \frac{1}{n-j} \sum_{i=1}^{n-j} C_{i,j} \left(\hat{\delta}_{i,j} - \delta_j \right)^2.$$

sont des estimateurs sans biais de σ_j^2 et de τ_j^2 .

Démonstration. □

Lemme 15. Pour $1 \leq j \leq n - 2$, les estimateurs :

$$\hat{\sigma}_j^2 := \frac{1}{n-j} \sum_{i=1}^{n-j+1} E_i \left(\frac{N_{i,j}}{E_i} - \hat{\lambda}_j \right)^2,$$

$$\hat{\tau}_j^2 := \frac{1}{n-j-1} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{D_{i,j+1}}{C_{i,j}} - \hat{\delta}_j \right)^2.$$

sont des estimateurs sans biais de σ_j^2 et de τ_j^2 .

Démonstration. Admis. □

5.3 Le cas particulier du nombre de sinistres au-dessus d'un seuil

En réassurance, on s'intéresse parfois uniquement aux sinistres au-dessus d'un seuil qui joue le rôle de franchise. En assurance, on est parfois amené à séparer les petits sinistres des grands sinistres, également via un seuil, et d'estimer le nombre de sinistres au-dessus de ce seuil.

Le modèle précédemment développé s'adapte parfaitement à ce cas. En particulier :

- Les variables aléatoires $(N_{i,j})_{1 \leq i,j \leq n}$ représentent le nombre de nouveaux sinistres qui dépassent un seuil fixé. Il peut s'agir de sinistres complètement nouveaux dont le coût estimé est supérieur au seuil, comme de sinistres déclarés antérieurement dont la valeur l'année précédente était sous le seuil.
- Les variables aléatoires $(D_{i,j})_{1 \leq i,j \leq n}$ représente le nombre de sinistre au dessus du seuil en année de développement $j - 1$ dont le coût estimé chute sous le seuil.

Nous avons à nouveau la matrice $(C_{i,j})_{1 \leq i,j \leq n}$ qui représente cette fois-ci le nombre de sinistres survenus en année i et au-dessus du seuil en année de développement j . Par construction nous avons cette fois-ci la contrainte supplémentaire $D_{i,j} \geq 0$.

Nous avons l'hypothèse suivante :

Hypothèse 4.

$$H2' \ N_{i,j} \mid \mathcal{F}_{i+j-2} \sim \mathcal{P}(\lambda_j E_i) \quad \text{et} \quad D_{i,j+1} \mid \mathcal{F}_{i+j-1} \sim \mathcal{B}(C_{i,j}, \delta_j)$$

Remarque 6. L'hypothèse $H2'$ implique les hypothèses $H2$ et $H3$.

Démonstration. Pour $H2$ il suffit d'appliquer l'espérance. Pour $H3$, il suffit d'appliquer la variance et de poser $\sigma_j^2 := \lambda_j$ et $\tau_j^2 := \delta_j(1 - \delta_j)$. □

Proposition 7. Sous $H2'$, les estimateurs précédemment définis :

$$\hat{\lambda}_j := \frac{\sum_{i=1}^{n-j+1} N_{i,j}}{\sum_{i=1}^{n-j+1} E_i},$$

$$\hat{\delta}_j := \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}.$$

sont également les estimateurs du maximum de vraisemblance.

Démonstration. **Étape 1.** Pour $1 \leq j \leq n$, on observe $(N_{i,j} \mid \mathcal{F}_{i+j-2})_{1 \leq i \leq n-j+1}$ qui suivent indépendamment la loi $\mathcal{P}(\lambda_j E_i)$. La log-vraisemblance, notée $\ell(\lambda_j)$, s'écrit :

$$\ell(\lambda_j) = - \sum_{i=1}^{n-j+1} \lambda_j E_i + \sum_{i=1}^{n-j+1} N_{i,j} \log(\lambda_j E_i) - \sum_{i=1}^{n-j+1} \log(N_{i,j}!)$$

$$\ell(\lambda_j) \propto - \sum_{i=1}^{n-j+1} \lambda_j E_i + \sum_{i=1}^{n-j+1} N_{i,j} \log(\lambda_j)$$

On a :

$$\ell'(\lambda_j) = - \sum_{i=1}^{n-j+1} E_i + \frac{1}{\lambda_j} \sum_{i=1}^{n-j+1} N_{i,j}.$$

On en déduit la condition du premier ordre :

$$\ell'(\hat{\lambda}_j) = 0 \iff \hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} N_{i,j}}{\sum_{i=1}^{n-j+1} E_i}.$$

De plus,

$$\ell''(\hat{\lambda}_j) = -\frac{1}{\hat{\lambda}_j^2} \sum_{i=1}^{n-j+1} N_{i,j} < 0.$$

Étape 2. Pour $1 \leq j \leq n-1$, on observe $(D_{i,j} \mid \mathcal{F}_{i+j-1})_{1 \leq i \leq n-j}$ qui suivent indépendamment la loi $\mathcal{B}(C_{i,j}, \delta_j)$. La log-vraisemblance, notée $\ell(\delta_j \mid \mathcal{B}_j)$, s'écrit :

$$\begin{aligned} \ell(\delta_j \mid \mathcal{B}_j) &= \sum_{i=1}^{n-j} \log \binom{C_{i,j}}{D_{i,j}} + \sum_{i=1}^{n-j} D_{i,j} \log(\delta_j) + \sum_{i=1}^{n-j} (C_{i,j} - D_{i,j}) \log(1 - \delta_j) \\ \ell(\delta_j \mid \mathcal{B}_j) &\propto \sum_{i=1}^{n-j} D_{i,j} \log(\delta_j) + \sum_{i=1}^{n-j} (C_{i,j} - D_{i,j}) \log(1 - \delta_j) \end{aligned}$$

On a :

$$\ell'(\delta_j \mid \mathcal{B}_j) = \frac{1}{\delta_j} \sum_{i=1}^{n-j} D_{i,j} - \frac{1}{1 - \delta_j} \sum_{i=1}^{n-j} (C_{i,j} - D_{i,j}).$$

On en déduit la condition du premier ordre :

$$\begin{aligned} \ell'(\hat{\delta}_j \mid \mathcal{B}_j) = 0 &\iff (1 - \hat{\delta}_j) \sum_{i=1}^{n-j} D_{i,j} - \hat{\delta}_j \sum_{i=1}^{n-j} (C_{i,j} - D_{i,j}) = 0 \\ &\iff \sum_{i=1}^{n-j} D_{i,j} - \hat{\delta}_j \sum_{i=1}^{n-j} C_{i,j} = 0 \\ &\iff \hat{\delta}_j := \frac{\sum_{i=1}^{n-j} D_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}} \end{aligned}$$

De plus,

$$\ell''(\hat{\delta}_j \mid \mathcal{B}_j) = -\frac{1}{\hat{\delta}_j^2} \sum_{i=1}^{n-j} D_{i,j} - \frac{1}{(1 - \hat{\delta}_j)^2} \sum_{i=1}^{n-j} (C_{i,j} - D_{i,j}) < 0$$

□

Proposition 8. Les estimateurs $\hat{\lambda}_j$ et $\hat{\delta}_j$ sont efficaces.

Démonstration. On calcule l'information de Fischer, notée $I(\lambda_j)$ associée à la log-vraisemblance $\ell(\lambda_j)$:

$$I(\lambda_j) = -\mathbb{E}[\ell''(\lambda_j) \mid \mathcal{B}_j] = \frac{1}{\lambda_j^2} \sum_{i=1}^{n-j+1} \lambda_j E_i = \frac{\sum_{i=1}^{n-j+1} E_i}{\lambda_j}.$$

Or, comme $\sigma_j^2 = \lambda_j$ par $H2'$, par le Lemme 12, on sait que :

$$\text{Var}(\hat{\lambda}_j \mid \mathcal{B}_j) = \frac{\sigma_j^2}{\sum_{i=1}^{n-j+1} E_i} = I^{-1}(\lambda_j).$$

L'estimateur $\hat{\lambda}_j$ est efficace.

On calcule l'information de Fischer, notée $I(\delta_j)$ associée à la log-vraisemblance $\ell(\delta_j)$:

$$I(\delta_j \mid \mathcal{B}_j) = -\mathbb{E}[\ell''(\delta_j \mid \mathcal{B}_j) \mid \mathcal{B}_j] = \frac{1}{\delta_j^2} \sum_{i=1}^{n-j} \delta_j C_{i,j} + \frac{1}{(1-\delta_j)^2} \sum_{i=1}^{n-j} (C_{i,j} - \delta_j C_{i,j}) = \frac{\sum_{i=1}^{n-j} C_{i,j}}{\delta_j(1-\delta_j)}.$$

Or, comme $\tau_j^2 = \delta_j(1-\delta_j)$ par $H2'$, par le Lemme 12, on sait que :

$$\text{Var}(\hat{\delta}_j \mid \mathcal{B}_j) = \frac{\tau_j^2}{\sum_{i=1}^{n-j} C_i} = I^{-1}(\delta_j \mid \mathcal{B}_j).$$

L'estimateur $\hat{\delta}_j$ est efficace. □

On peut obtenir la loi des $C_{i,j}$, en particulier de $C_{i,n}$, conditionnellement à \mathcal{F}_n . Nous allons utiliser le lemme suivant.

Lemme 16. *Soit N une variable aléatoire de loi $\mathcal{P}(\lambda)$ avec $\lambda > 0$ et D une variable aléatoire telle que $D \mid N \sim \mathcal{B}(N, p)$ avec $p > 0$.*

Alors

$$N - D \sim \mathcal{P}(\lambda(1-p)).$$

Démonstration. Soit $k \in \mathbb{N}$.

$$\begin{aligned}
\mathbb{P}(N - D = k) &= \mathbb{P}\left(\{N - D = k\} \cap \left\{\bigcup_{n \in \mathbb{N}} \{N = n\}\right\}\right) \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(\{N - D = k\} \cap \{N = n\}) \\
&= \sum_{n \in \mathbb{N}} \mathbb{P}(\{D = k\} \mid \{N = n\}) \mathbb{P}(N = n) \\
&= \sum_{n \geq k} \frac{n!}{k!(n-k)!} p^{n-k} (1-p)^k \frac{e^{-\lambda} \lambda^n}{n!} \\
&= \frac{e^{-\lambda} (1-p)^k}{k!} \sum_{n \geq 0} \frac{p^n \lambda^{n+k}}{n!} \\
&= \frac{e^{-\lambda} [\lambda(1-p)]^k}{k!} \sum_{n \geq 0} \frac{(p\lambda)^n}{n!} \\
&= \frac{e^{-\lambda(1-p)} [\lambda(1-p)]^k}{k!}.
\end{aligned}$$

□

Lemme 17. Pour tout $1 \leq i, j \leq n$, on a

$$C_{i,j} \sim \mathcal{P}(\lambda_{i,j}),$$

avec

$$\lambda_{i,j} := \sum_{k=1}^j \lambda_k E_i \left(\prod_{\ell=k}^{j-1} (1 - \delta_\ell) \right).$$

Démonstration. On raisonne par récurrence. On fixe $1 \leq i \leq n$. On a $C_{i,1} = N_{i,1} \sim \mathcal{P}(\lambda_1 E_i)$. On suppose par hypothèse de récurrence que $C_{i,j}$ suit une loi de Poisson de paramètre $\lambda_{i,j} := \sum_{k=1}^j \lambda_k E_i \left(\prod_{\ell=k}^{j-1} (1 - \delta_\ell) \right)$

$$C_{i,j+1} = C_{i,j} + N_{i,j+1} - D_{i,j+1}.$$

Sous H_2' , par le lemme précédent,

$$C_{i,j} - D_{i,j+1} \sim \mathcal{P}(\lambda_{i,j}(1 - \delta_j)),$$

puis

$$C_{i,j+1} \sim \mathcal{P}(\lambda_{j+1} E_i + \lambda_{i,j}(1 - \delta_j)).$$

□

Lemme 18. Sous H_1 et H_2' , nous avons :

$$\hat{\lambda}_j \sim \frac{\mathcal{P}\left(\lambda_j \sum_{i=1}^{n-j+1} E_i\right)}{\sum_{i=1}^{n-j+1} E_i}$$

$$\hat{\delta}_j \sim \frac{\mathcal{B}\left(\sum_{i=1}^{n-j} C_{i,j}, \delta_j\right)}{\sum_{i=1}^{n-j} C_{i,j}}$$

Démonstration. Par définition, on a :

$$\hat{\lambda}_j | \mathcal{B}_{j-1} := \frac{\sum_{i=1}^{n-j+1} N_{i,j}}{\sum_{i=1}^{n-j+1} E_i},$$

$$\hat{\delta}_j | \mathcal{B}_j := \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}.$$

Par $H1$, les $N_{i,j}$ sont indépendants entre eux, tout comme les $D_{i,j+1}$. Par $H2'$, comme $N_{i,j} | \mathcal{B}_j \sim \mathcal{P}(\lambda_j E_i)$ et $D_{i,j+1} \sim \mathcal{B}(C_{i,j}, \delta_j)$, on obtient le résultat. \square

Cette fois-ci, on ne définit plus des réserves, mais le nombre estimé de nouveaux sinistres qui seront au-dessus du seuil : $\hat{M}_i := \hat{C}_{i,n} - C_{i,n-i+1}$.

5.4 Exemple

Nous avons le triangle cumulé suivant (tiré de [3]) auquel on associe l'exposition.

i	E_i	$C_{i,1}$	$C_{i,2}$	$C_{i,3}$	$C_{i,4}$	$C_{i,5}$	$C_{i,6}$	$C_{i,7}$
1	10224	7.5	28.9	52.6	84.5	80.1	76.9	79.5
2	12752	1.6	14.8	32.1	39.6	55	60	
3	14875	13.8	42.4	36.3	53.3	96.5		
4	17365	2.9	14	32.5	46.9			
5	19410	2.9	9.8	52.7				
6	17617	1.9	29.4					
7	18129	19.1						

Celui-ci se décompose en la matrice N :

i	$N_{i,1}$	$N_{i,2}$	$N_{i,3}$	$N_{i,4}$	$N_{i,5}$	$N_{i,6}$	$N_{i,7}$
1	7.5	18.3	28.5	23.4	18.6	0.7	5.1
2	1.6	12.6	18.2	16.1	14	10.6	
3	13.8	22.7	4	12.4	12.1		
4	2.9	9.7	16.4	11.6			
5	2.9	6.9	37.1				
6	1.9	27.5					
7	19.1						

et D :

i	$D_{i,1}$	$D_{i,2}$	$D_{i,3}$	$D_{i,4}$	$D_{i,5}$	$D_{i,6}$	$D_{i,7}$
1	0	-3.1	4.8	-8.5	23	3.9	2.5
2	0	-0.6	0.9	8.6	-1.4	5.6	
3	0	-5.9	10.1	-4.6	-31.1		
4	0	-1.4	-2.1	-2.8			
5	0	0	-5.8				
6	0	0					
7	0						

Si les triangles ci-dessus sont importés dans des matrices ou des data.frame qui s'appellent respectivement C, N et D dans le langage R, et l'exposition dans un vecteur E, alors le code suivant permet de calculer $(\hat{\lambda}_j)_{1 \leq j \leq n}$ et $(\hat{\delta}_j)_{1 \leq j \leq n-1}$.

```

lambda <- numeric(n)
for(j in 1L:n)
  lambda[j] <- sum(N[1L:(n-j+1L), j])/sum(E[1L:(n-j+1L)])

delta <- numeric(n-1L)
for(j in 1L:(n-1L))
  delta[j] <- sum(D[1L:(n-j), j + 1L])/sum(C[1L:(n-j), j])

Le code suivant permet de calculer  $(\hat{\sigma}_j^2)_{1 \leq j \leq n}$  et  $(\hat{\tau}_j^2)_{1 \leq j \leq n-1}$ .

s2 <- numeric(n)
for(j in 1L:(n-1L))
  s2[j] <- sum( E[1L:(n-j+1)] * (N[1L:(n-j+1), j]/E[1L:(n-j+1)]
    - lambda[j])^2 )/(n - j)
s2[n] <- 0

tau2 <- numeric(n-1L)
for(j in 1L:(n-2L))
  tau2[j] <- sum( C[1L:(n-j), j] * (D[1L:(n-j), j + 1L]/C[1L:(n-j), j]
    - delta[j])^2 )/(n - j - 1L)
tau2[n-1L] <- 0

```

On obtient

j	1	2	3	4	5	6	7
$\hat{\lambda}_j \times 10^{-3}$	0.450	1.059	1.396	1.150	1.181	0.492	0.499
$\hat{\delta}_j^2$	-0.359	0.072	-0.048	-0.054	0.070	0.033	
$\hat{\sigma}_j$	0.0538	0.0737	0.1089	0.0795	0.0560	0.0575	0.0000
$\hat{\tau}_j^2$	0.3874	1.2686	1.1768	3.4603	0.3034	0.0000	

6 Méthode du bootstrap

La méthode du bootstrap s'appuie sur un rééchantillonnage de l'erreur. Il permet ensuite de calculer, par exemple, des intervalles de confiance sur différents estimateurs.

Le cas le plus simple est celui de l'estimation d'un intervalle de confiance pour la moyenne empirique. Si on observe : (X_1, \dots, X_n) , un échantillon i.i.d. de L^2 , en notant $\mu := \mathbb{E}(X_1)$ et $\sigma := \sqrt{\text{Var}(X_1)}$, on sait par le théorème central limite que

$$\sqrt{n} \frac{(\bar{X}_n - \mu)}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

ce qui permet d'en déduire un intervalle de confiance **asymptotique**. L'idée du bootstrap est d'utiliser la variabilité intrinsèque de l'échantillon afin d'en construire une distribution estimée de l'estimateur autour de μ .

Dans ce cas, on rééchantillonne (X_1, \dots, X_n) , c'est à dire qu'on introduit :

$$(X_1^m, \dots, X_n^m) \stackrel{i.i.d.}{\sim} \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \quad 1 \leq m \leq N.$$

On construit ainsi N nouveaux échantillons de taille n où chaque X_i^m suit de manière indépendante la loi uniforme sur l'échantillon observé (X_1, \dots, X_n) . On calcule ensuite les moyennes empiriques associées $(\hat{\mu}^m)_{1 \leq m \leq N}$ ce qui donne une distribution de l'estimateur.

Par exemple, si on observe $(5, 5, 7, 4, 11, 4, 10, 8, 6, 7, 12, 5)$, on a $\hat{\mu} = 7$ et $\hat{\sigma} \approx 2.73$. On en déduit sous le TCL l'intervalle au niveau 95% : $[5.46, 8.54]$.

En revanche, si on applique la procédure du bootstrap un grand nombre de fois, et qu'on prend l'intervalle de confiance sur la distribution de $(\hat{\mu}^m)_{1 \leq m \leq N}$, on obtient au niveau 95% : $[5.58, 8.50]$.

Dans notre cas, il s'agit de *repérer* ce qui est (ou sera supposé) i.i.d. Nous allons dans un premier temps regarder le cas du modèle de Mack puis dans un second temps celui de Schnieper.

6.1 Modèle de Mack

Dans le modèle de Mack, nous avons comme hypothèse : $\mathbb{E}(C_{i,j} | \mathcal{B}_j) = f_j C_{i,j}$ et $\text{Var}(C_{i,j} | \mathcal{B}_j) = \sigma_j^2 C_{i,j}$.

On définit le résidu de Pearson de la manière suivante :

$$r_{i,j} := \frac{\frac{C_{i,j+1}}{C_{i,j}} - f_j}{\frac{\sigma_j}{\sqrt{C_{i,j}}}} = \frac{C_{i,j+1} - f_j C_{i,j}}{\sigma_j \sqrt{C_{i,j}}}.$$

On va faire l'hypothèse supplémentaire que $r_{i,j}$ a la même loi pour $1 \leq i \leq n$ et $1 \leq j \leq n-1$. Comme les paramètres f_j et σ_j sont inconnus, on les remplace par leurs estimateurs, on définit pour $i + j + 1 \leq n + 1$:

$$\hat{r}_{i,j} := \frac{C_{i,j+1} - \hat{f}_j C_{i,j}}{\hat{\sigma}_j \sqrt{C_{i,j}}}.$$

C'est l'échantillon sur lequel nous allons appliquer le bootstrap. On simule $(\hat{r}_{i,j}^m)_{1 \leq m \leq N}$ en tirant uniformément et indépendamment dans l'échantillon $(\hat{r}_{i,j})_{i+j \leq n}$. On recalcule le triangle supérieur avec la relation :

$$\begin{aligned} C_{i,1}^m &:= C_{i,1}, \\ C_{i,j+1}^m &:= \hat{f}_j^m C_{i,j}^m + \hat{r}_{i,j}^m \hat{\sigma}_j^m \sqrt{C_{i,j}^m}. \end{aligned}$$

Cela permet d'en déduire un nouveau triangle sur lequel on estime à nouveau les paramètres f_j , on note \hat{f}_j^m les estimateurs correspondants. Ensuite, pour la partie inconnue du triangle, on la simule selon :

$$\begin{aligned} C_{i,n-i+1}^m &:= C_{i,n-i+1}, \\ C_{i,j+1}^m &:= \hat{f}_j^m C_{i,j}^m + \hat{r}_{i,j}^m \hat{\sigma}_j^m \sqrt{C_{i,j}^m}. \end{aligned}$$

On peut ensuite calculer :

$$R_i^m = C_{i,n}^m - C_{i,n-i+1}^m, \quad 1 \leq i \leq n, \quad 1 \leq m \leq N.$$

Ces simulations tiennent à la fois compte de l'erreur d'estimation des paramètres (première étape de reconstruction du triangle où on estime à nouveau les paramètres) et du bruit (seconde étape où on simule la partie inconnue du triangle).

6.2 Modèle de Schnieper

Nous allons cette fois-ci définir des résidus pour chacun des triangles. Le triangle des nouveaux sinistres, nous avons comme hypothèse : $\mathbb{E}(N_{i,j} \mid \mathcal{B}_{j-1}) = \lambda_j E_i$ et $\text{Var}(N_{i,j} \mid \mathcal{B}_{j-1}) = \sigma_j^2 E_i$.

On définit le résidu de Pearson de la manière suivante :

$$n_{i,j} := \frac{\frac{N_{i,j}}{E_i} - \lambda_j}{\frac{\sigma_j}{\sqrt{E_i}}} = \frac{N_{i,j} - \lambda_j E_i}{\sigma_j \sqrt{E_i}}.$$

De manière analogue, compte tenu des hypothèse sur $D_{i,j+1}$, on définit :

$$d_{i,j} := \frac{\frac{D_{i,j+1}}{C_{i,j}} - \delta_j}{\frac{\tau_j}{\sqrt{C_{i,j}}}} = \frac{D_{i,j+1} - \delta_j C_{i,j}}{\tau_j \sqrt{C_{i,j}}}.$$

On va faire l'hypothèse supplémentaire que $n_{i,j}$ a la même loi pour $1 \leq i \leq n$ et $1 \leq j \leq n$. Comme les paramètres λ_j et σ_j sont inconnus, on les remplace par leurs estimateurs, on définit pour $i + j \leq n + 1$:

$$\hat{n}_{i,j} := \frac{N_{i,j} - \hat{\lambda}_j E_i}{\hat{\sigma}_j \sqrt{E_i}}.$$

Et de même :

$$\hat{d}_{i,j} := \frac{D_{i,j+1} - \hat{\delta}_j C_{i,j}}{\hat{\tau}_j \sqrt{C_{i,j}}}.$$

On simule $(\hat{n}_{i,j}^m)_{1 \leq m \leq N}$ et $(\hat{d}_{i,j}^m)_{1 \leq m \leq N}$ en tirant uniformément et indépendamment respectivement dans l'échantillon $(\hat{n}_{i,j})_{i+j \leq n}$ et $(\hat{d}_{i,j})_{i+j \leq n+1}$. On recalcule le triangle supérieur avec la relation :

$$\begin{aligned} C_{i,1}^m &:= \hat{\lambda}_1 E_i + \hat{n}_{i,j}^m \hat{\sigma}_1 \sqrt{E_i}, \\ C_{i,j+1}^m &:= \hat{\lambda}_{j+1} E_i + \hat{n}_{i,j}^m \hat{\sigma}_{j+1} \sqrt{E_i} + (1 - \hat{\delta}_j) C_{i,j}^m + \hat{d}_{i,j}^m \hat{\tau}_j \sqrt{C_{i,j}^m}. \end{aligned}$$

Cela permet d'en déduire un nouveau triangle sur lequel on estime à nouveau les paramètres λ_j et δ_j . On note respectivement $\hat{\lambda}_j^m$ et $\hat{\delta}_j^m$ les estimateurs correspondants. Ensuite, pour la partie inconnue du triangle, on la simule selon :

$$\begin{aligned} C_{i,n-i+1}^m &:= C_{i,n-i+1}, \\ C_{i,j+1}^m &:= \hat{\lambda}_{j+1}^m E_i + \hat{n}_{i,j}^m \hat{\sigma}_{j+1}^m \sqrt{E_i} + (1 - \hat{\delta}_j^m) C_{i,j}^m + \hat{d}_{i,j}^m \hat{\tau}_j^m \sqrt{C_{i,j}^m}. \end{aligned}$$

On peut ensuite calculer :

$$R_i^m = C_{i,n}^m - C_{i,n-i+1}, \quad 1 \leq i \leq n, \quad 1 \leq m \leq N.$$

Ces simulations tiennent à la fois compte de l'erreur d'estimation des paramètres (première étape de reconstruction du triangle où on estime à nouveau les paramètres) et du bruit (seconde étape où on simule la partie inconnue du triangle).

Pour le cas des **nombres** au-dessus d'un seuil, sous les hypothèses plus fortes de distribution, nous ne passons plus par les réidus. Nous simulons directement les valeurs. Plus précisément, pour le triangle supérieur, nous simulons :

$$\begin{aligned} C_{i,1}^m &\sim \mathcal{P}(\hat{\lambda}_1 E_i), \\ C_{i,j+1}^m &\sim C_{i,j}^m + \mathcal{P}(\hat{\lambda}_{j+1} E_i) - \mathcal{B}(C_{i,j}^m, \hat{\delta}_j). \end{aligned}$$

Cela permet à nouveau d'en déduire un nouveau triangle sur lequel on estime à nouveau les paramètres λ_j et δ_j . On note à nouveau respectivement $\hat{\lambda}_j^m$ et $\hat{\delta}_j^m$ les estimateurs correspondants. Ensuite, pour la partie inconnue du triangle, on la simule selon :

$$\begin{aligned} C_{i,n-i+1}^m &:= C_{i,n-i+1}, \\ C_{i,j+1}^m &:= C_{i,j}^m + \mathcal{P}(\hat{\lambda}_{j+1}^m E_i) - \mathcal{B}(C_{i,j}^m, \hat{\delta}_j^m). \end{aligned}$$

On peut ensuite calculer le nombre de sinistres manquant au-dessus du seuil :

$$M_i^m = C_{i,n}^m - C_{i,n-i+1}, \quad 1 \leq i \leq n, \quad 1 \leq m \leq N.$$

7 Prise en compte de l'inflation

L'inflation joue un rôle important. Les sinistres qui ont lieu dans une année i sont réglés à différents moments. De plus, le coût des sinistres sur chaque année est touché par l'inflation. Il convient donc de déflater le triangle, afin de travailler en **euros constants**.

Si on suppose que $X_{i,j}$ (le coût incrémental) correspond au nouveau coût réel ou la charge estimée à venir, évaluée en année $i + j$, et que $(I_m)_{m \geq 1}$ est l'indice des prix associé au secteur de notre triangle, alors on introduit :

$$X'_{i,j} := X_{i,j} \frac{I_{n+1}}{I_{i+j}}.$$

On en déduit le nouveau triangle cumulé dans le cas de Chain Ladder Mack (et de manière analogue dans le cas de Schnieper). On fait ensuite les estimations sur les triangles déflatés. Enfin, pour compléter le triangle, on applique l'inflation en multipliant par $\frac{I_{i+j}}{I_{n+1}}$.

Références

- [1] Nicolas Baradel. Cours de théorie du risque. https://nicolasbaradel.fr/enseignement/ressources/cours_theorie_du_risque.pdf.
- [2] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin : The Journal of the IAA*, 23(2) :213–225, 1993.
- [3] R Schnieper. Separating true ibnr and ibner claims 1. *ASTIN Bulletin : The Journal of the IAA*, 21(1) :111–127, 1991.