

ASSURANCE NON-VIE



Année universitaire 2004-2005

Le modèle collectif

Pierre-E. THEROND

ptherond@jwa.fr

1. Introduction	2
2. Lois composés	2
2.1. Loi de Poisson composée	2
2.2. Loi de Poisson mélange.....	3
2.3. Loi de Poisson mélange composée.....	4
3. Du modèle individuel au modèle collectif.....	5
3.1. Le modèle individuel.....	5
3.2. Le modèle collectif.....	6
3.3. Approximation collective du modèle individuel.....	6
4. Discrétisation dans le modèle collectif.....	8
4.1. Discrétisation pour calculer une probabilité de ruine.....	9
4.2. Discrétisation pour calculer une prime stop-loss	9
5. Algorithme de Panjer	10
5.1. Calcul direct de la fonction de répartition discrétisée	10
5.2. Algorithme de Panjer	10
5.3. Calcul des primes stop-loss	11
6. Méthodes d'approximation de la loi de S^{coll}	12
6.1. Fast Fourier Transform	13
6.2. Approximation Gamma de Bowers.....	14
Bibliographie.....	16
Exercices.....	16

1. Introduction

Lorsque l'assureur s'intéresse à la charge globale des sinistres engendrés par un portefeuille, il lui est souvent impossible en pratique d'analyser le risque police par police alors qu'il est plus vraisemblable de l'analyser au niveau du portefeuille de contrats par le biais de la charge globale de sinistres de ce portefeuille $S = \sum_{i=1}^n S_i$ où n est le nombre de polices et S_i la charge de sinistres provenant de la police n° i . Le modèle collectif est un outil qui nous permet d'approcher la loi de S .

2. Loïs composés

Si N désigne le nombre de sinistres survenus sur la période considérée et $X_1, X_2, \dots$ les montants de ceux-ci, le montant de S la charge globale de sinistres sur la période est donné par :

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

Dans la suite, nous supposons que les variables aléatoires $X_1, X_2, \dots, N$ sont mutuellement indépendantes et que $X_1, X_2, \dots$ est un i.i.d. de même loi que X .

L'objet de ce paragraphe est d'expliciter la loi de S en fonction des lois de X et de N . On se penchera en particulier sur les cas où N suit une loi de Poisson ou une loi de Poisson mélange.

2.1. Loi de Poisson composée

Définition 1 : Si $N \sim P(\lambda)$ et si les $X_1, X_2, \dots$ admettent F comme fonction de répartition, S suit une loi de Poisson composée $PC(\lambda, F)$.

Dans ce cas, on a :

$$G(s) = \mathbf{Pr}[S \leq s] = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} F^{*(k)}(s), s \geq 0,$$

$$\mathbf{E}[S] = \lambda \mathbf{E}[X],$$

$$\mathbf{Var}[S] = \lambda \mathbf{Var}[X] + \lambda \mathbf{E}^2[X] = \lambda \mathbf{E}[X^2],$$

et

$$L_S(t) = \exp\{\lambda(L_X(t) - 1)\}, t \geq 0.$$

Il est rappelé que L_X la transformée Laplace d'une variable aléatoire X est donnée par

$$\forall t \geq 0, L_X(t) = \mathbf{E}[\exp\{-tX\}].$$

2.2. Loi de Poisson mélange

Définition 2 : N suit une loi de Poisson mélange $PM(\lambda, \Theta)$ de niveau de risque Θ lorsque

$$\Pr[N = k] = \mathbf{E} \left[\exp(-\lambda\Theta) \frac{(\lambda\Theta)^k}{k!} \right] \text{ où } \Theta \text{ satisfait la contrainte } \mathbf{E}[\Theta] = 1.$$

La variable aléatoire Θ pourra en particulier modéliser l'hétérogénéité du portefeuille en fréquence. Pour illustrer cela supposons que le portefeuille est constitué de deux types de risques : des « bons » risques et des « mauvais » risques qui sont en proportions égales dans le portefeuille. Supposons également que les bons risques aient en moyenne trois fois moins de sinistres que les « mauvais » risques.

Formellement, on a :

$$\begin{cases} \Pr[\Theta = \theta_1] = \Pr[\Theta = \theta_2] = \frac{1}{2} \\ \mathbf{E}[\Theta | \Theta = \theta_1] = \frac{1}{3} \mathbf{E}[\Theta | \Theta = \theta_2]. \end{cases}$$

Comme

$$\mathbf{E}[\Theta] = \Pr[\Theta = \theta_1] * \mathbf{E}[\Theta | \Theta = \theta_1] + \Pr[\Theta = \theta_2] * \mathbf{E}[\Theta | \Theta = \theta_2] = 1,$$

il vient

$$\frac{1}{2} * \{\mathbf{E}[\Theta | \Theta = \theta_1] + \mathbf{E}[\Theta | \Theta = \theta_2]\} = 1.$$

Donc

$$\begin{cases} \mathbf{E}[\Theta | \Theta = \theta_1] = \theta_1 = \frac{1}{2} \\ \mathbf{E}[\Theta | \Theta = \theta_2] = \theta_2 = \frac{3}{2}. \end{cases}$$

La probabilité qu'une police dont on ne sait pas s'il s'agit d'un « bon » ou d'un « mauvais » risque produise k sinistres est alors donnée par :

$$\Pr[N = k] = \frac{1}{2} \exp\left(-\frac{\lambda}{2}\right) \frac{(\lambda/2)^k}{k!} + \frac{1}{2} \exp\left(-\frac{3\lambda}{2}\right) \frac{(3\lambda/2)^k}{k!}.$$

Proposition 1 : (Théorème de Shaked) Si $N \sim PM(\lambda, \Theta)$ alors il existe deux valeurs entières $0 \leq k_0 < k_1$ telles que

$$\Pr[N = k] \geq \exp(-\lambda) \frac{\lambda^k}{k!} \text{ pour } k = 0, 1, \dots, k_0,$$

$$\Pr[N = k] \leq \exp(-\lambda) \frac{\lambda^k}{k!} \text{ pour } k = k_0 + 1, \dots, k_1,$$

$$\Pr[N = k] \geq \exp(-\lambda) \frac{\lambda^k}{k!} \text{ pour } k \geq k_1 + 1.$$

Démonstration : Cf. DENUIT et CHARPENTIER [2004]. □

Ce théorème nous indique que la superposition d'une erreur Θ à la moyenne λ a pour effet d'augmenter les masses de probabilité en 0 et pour les grandes valeurs par rapport à la loi de Poisson de même moyenne λ .

2.3. Loi de Poisson mélange composée

Définition 3 : Si $N \sim PM(\lambda, \Theta)$ et si les $X_1, X_2, \dots$ admettent F comme fonction de répartition, S suit une loi de Poisson mélange composée $PMC(\lambda, \Theta, F)$.

Si $S \sim PMC(\lambda, \Theta, F)$, on a

$$\mathbf{Var}[S] = \lambda \mathbf{E}[X^2] + \lambda^2 \mathbf{E}[X] \mathbf{Var}[\Theta].$$

Cette expression peut se réécrire

$$\mathbf{Var}[S] = \mathbf{E}^2[X] \mathbf{Var}[P(\lambda)] + \lambda \mathbf{Var}[X] + \lambda^2 \mathbf{E}[X] \mathbf{Var}[\Theta].$$

Le premier terme du membre de droite de cette égalité peut se voir comme $\mathbf{Var}[P(\lambda) \mathbf{E}[X]]$ c'est à dire comme la variance du montant global de sinistre si les sinistres étaient tous de montant égal à leur moyenne et comme si leur nombre était une loi de Poisson. Le deuxième terme peut se réécrire¹ $\mathbf{Var}\left[\sum_{i=1}^{\lambda} X_i\right]$ et peut se voir comme la contribution des montants de sinistres à la variabilité totale. Le troisième terme peut donc être vu comme la variabilité ajoutée par la mélangeante.

Remarquons que lorsque $\lambda \rightarrow +\infty$, $CV(S) = \sqrt{\mathbf{Var}[S]}/\mathbf{E}[S] \rightarrow \sqrt{\mathbf{Var}[\Theta]}$. Ce qui signifie que lorsque le nombre moyen de sinistres (et donc le nombre de polices) devient très grand, le principal risque dans le portefeuille provient de l'hétérogénéité de celui-ci.

Lorsqu'il n'y a pas d'hétérogénéité, *i. e.* lorsque $\Theta = 1$, $S \sim PC(\lambda, F)$ et $CV(S) \rightarrow 0$ lorsque $\lambda \rightarrow +\infty$. La loi des grands nombres nous indique que

$$S \xrightarrow{\text{Pr}} \mathbf{E}[S] \text{ lorsque } \lambda \rightarrow +\infty.$$

Plus généralement, si $S \sim PMC(\lambda, \Theta, F)$ alors

$$\frac{S}{\mathbf{E}[S]} \xrightarrow{\text{Loi}} \Theta \text{ lorsque } \lambda \rightarrow +\infty.$$

Le ratio sinistres à primes a donc pour loi limite la mélangeante décrivant l'hétérogénéité du risque en fréquence.

¹ Cette écriture est formelle dans la mesure où λ n'est pas nécessairement entier.

3. Du modèle individuel au modèle collectif

La charge globale de sinistres peut s'écrire comme la somme, sur le nombre de polices, du montant de sinistre total engendré par chaque police ou encore comme la somme, sur le nombre de sinistres, des montants de chaque sinistre. On appelle modèle individuel la première approche et modèle collectif la seconde.

3.1. Le modèle individuel

Le modèle individuel revient à raisonner police par police. Notons n le nombre de polices dans le portefeuille, la charge globale causée par les sinistres touchant ces sinistres est donnée par :

$$S^{ind} = \sum_{i=1}^n S_i,$$

où S_i désigne la charge totale des sinistres ayant atteint la police i sur la période considérée.

Considérons à présent q_i la probabilité que la police i produise au moins un sinistre sur la période et p_i celle qu'elle n'en produise aucun. Si $F_i(x) = \mathbf{Pr}[S_i \leq x]$ désigne la fonction de répartition de S_i , on a donc : $F_i(0) = p_i < 1$.

Si $B_i(x) = \mathbf{Pr}[S_i \leq x | S_i > 0]$ est la fonction de répartition de la charge des sinistres relative à la police i sachant que celle-ci a produit au moins un sinistre, on a

$$F_i(x) = p_i \mathbf{1}_{x \geq 0} + q_i B_i(x), \quad x > 0$$

et

$$B_i(x) = \frac{F_i(x) - F_i(0)}{1 - F_i(0)}, \quad x > 0.$$

Les deux premiers moments de la charge globale de sinistre s'expriment aisément en fonction des deux premiers moments (lorsqu'ils existent) des montants de sinistres par police :

$$\begin{cases} \mathbf{E}[S^{ind}] = \sum_{i=1}^n \mathbf{E}[S_i] \\ \mathbf{Var}[S^{ind}] = \sum_{i=1}^n \mathbf{Var}[S_i]. \end{cases}$$

Le montant de sinistre S_i engendré par la police i peut s'exprimer sous la forme

$$S_i = I_i Y_i,$$

où I_i vaut 1 si la police i a été touchée par au moins un sinistre, 0 sinon ; et où Y_i représente la somme des montants de sinistres qui ont touché la police i si au moins un sinistre est survenu.

Les variables aléatoires $I_1, I_2, \dots, I_n, Y_1, Y_2, \dots, Y_n$ sont supposées mutuellement indépendantes. Sous cette hypothèse, le montant de la prime pure est donné par

$$\mathbf{E}[S_i] = \mathbf{E}[I_i] \mathbf{E}[Y_i].$$

Sous l'hypothèse d'indépendance, en notant G^{ind} la fonction de répartition de S^{ind} , on a

$$G^{ind} = F_1 * F_2 * \dots * F_n.$$

Cette fonction de répartition est toutefois difficile à calculer. Le passage au modèle collectif va permettre de l'approcher.

3.2. Le modèle collectif

Le modèle collectif ne considère plus les polices individuellement mais le portefeuille dans son ensemble : la charge globale de sinistres va être exprimée en fonction du montant de chaque sinistre et non plus en fonction du montant total (éventuellement nul) des sinistres générés par chaque police. Ainsi notons N le nombre de sinistres touchant le portefeuille sur la période. Les montants de ces sinistres sont des variables aléatoires $X_1, X_2, \dots$ que l'on supposera dans la suite i.i.d. de fonction de répartition F et indépendantes de N .

La charge totale de sinistres générée par le portefeuille est donnée par

$$S^{coll} = \sum_{i=1}^N X_i.$$

3.3. Approximation collective du modèle individuel

Pour $t > 0$, la transformée de Laplace de S^{ind} est donnée par

$$L_{S^{ind}}(t) = \prod_{i=1}^n L_{S_i}(t) = \prod_{i=1}^n (p_i + q_i L_{Y_i}(t)).$$

On rappelle que $Y_i = [S_i | S_i > 0]$. De plus

$$p_i + q_i = 1 \Rightarrow L_{S^{ind}}(t) = \prod_{i=1}^n \{1 + q_i (L_{Y_i}(t) - 1)\}.$$

En passant au logarithme népérien, on obtient

$$\begin{aligned} \ln L_{S^{ind}}(t) &= \sum_{i=1}^n \ln \{1 + q_i (L_{Y_i}(t) - 1)\} \\ &= \sum_{i=1}^n \sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} \{q_i (L_{Y_i}(t) - 1)\}^k. \end{aligned}$$

En limitant le développement au degré 1, on obtient l'approximation

$$\ln L_{S^{ind}}(t) \approx \sum_{i=1}^n q_i (L_{Y_i}(t) - 1) = \lambda \{L_X(t) - 1\},$$

$$\text{où } \lambda = \sum_{i=1}^n q_i \text{ et } L_X(t) = \frac{1}{\lambda} \sum_{i=1}^n q_i L_{Y_i}(t).$$

Ainsi $\ln L_{S^{ind}}(t) \approx$ au $\ln$ de la transformée de Laplace d'une loi de Poisson composée $PC(\lambda, F)$ (cf. paragraphe 2.1).

En approximant la fonction de répartition F_i de S_i par $\tilde{F}_i$ définie par :

$$\tilde{F}_i(x) = \exp(-\lambda_i) \sum_{k=0}^{+\infty} \frac{\lambda_i^k}{k!} B_i^{*(k)}(x),$$

on peut, pour chaque police i , approximer S_i par $\tilde{S}_i \sim PC(\lambda_i, B_i)$ donné par :

$$\tilde{S}_i = \sum_{k=1}^{N_i} Y_{i,k}.$$

Il vient alors la charge totale dans le modèle collectif S^{coll} qui peut s'écrire

$$S^{coll} = \sum_{i=1}^n \tilde{S}_i = \sum_{k=1}^N X_i \sim PC(\lambda, F).$$

La fonction de répartition G^{coll} de S^{coll} est alors donnée par

$$G^{coll}(x) = \exp(-\lambda) \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} F^{*(k)}(x),$$

$$\text{où } \lambda = \sum_{i=1}^n \lambda_i \text{ et } F(x) = \frac{1}{\lambda} \sum_{i=1}^n \lambda_i B_i(x).$$

3.3.1. Choix des paramètres du modèle collectif

La mise en œuvre de l'approximation collective requiert de choisir les paramètres λ_i .

Le choix le plus naturel pour les λ_i consiste à prendre $\lambda_i = q_i$. Ce choix permet d'avoir l'égalité entre le nombre de polices produisant des sinistres dans le modèle individuel et le nombre moyen de sinistres dans le modèle collectif. En effet, avec ce choix

$$\sum_{i=1}^n q_i = \sum_{i=1}^n \lambda_i = \lambda = \mathbf{E}[N].$$

Ce choix aboutit par ailleurs à la même prime pure dans les deux modèles puisque $\mathbf{E}[S^{ind}] = \mathbf{E}[S^{coll}]$. En revanche, le modèle collectif surestime la variance de la charge globale de sinistres car $\mathbf{Var}[S_i] \leq \mathbf{Var}[\tilde{S}_i]$. Avec ces paramètres, le modèle collectif est donc prudent puisqu'il conduit à un risque supérieur à celui du modèle individuel.

Il est également fréquent d'utiliser $\lambda_i = -\ln(1 - q_i)$. Ce choix permet de ne pas léser les bons risques (les polices qui ne produisent pas de sinistre) puisque :

$$\begin{aligned}\lambda_i = -\ln(1 - q_i) &\Leftrightarrow \exp(-\lambda_i) = p_i \\ &\Leftrightarrow \mathbf{Pr}[N_i = 0] = \mathbf{Pr}[S_i = 0].\end{aligned}$$

Ce choix apparaît prudent puisque, sur l'ensemble du portefeuille, il conduit à une prime pure prudente : $\mathbf{E}[S^{ind}] \leq \mathbf{E}[S^{coll}]$.

3.3.2. Erreur d'approximation

Pour tout $x \geq 0$, on démontre (cf. DENUIT et CHARPENTIER [2004] pour le détail des calculs) que :

$$\begin{cases} \sum_{i=1}^n \min\{0, p_i - \exp(-\lambda_i)\} \leq \mathbf{Pr}[S^{ind} \leq x] - \mathbf{Pr}[S^{coll} \leq x] = G^{ind}(x) - G^{coll}(x) \\ G^{ind}(x) - G^{coll}(x) \leq \sum_{i=1}^n \{p_i - \exp(-\lambda_i) + \max\{q_i - \lambda_i \exp(-\lambda_i), 0\}\}. \end{cases}$$

Si $\lambda_i = q_i$, on a : $\forall x \in \mathbf{R}_+, |G^{ind}(x) - G^{coll}(x)| \leq \frac{1}{2} \sum_{i=1}^n q_i^2$.

Si $\lambda_i = -\ln p_i$, on a : $\forall x \in \mathbf{R}_+, 0 \leq G^{ind}(x) - G^{coll}(x) \leq \frac{1}{2} \sum_{i=1}^n (\ln p_i)^2$.

4. Discrétisation dans le modèle collectif

En pratique, le calcul de la fonction de répartition de S^{coll} va passer par celui des produits de convolution $F^{*(k)}$. F n'étant généralement pas stable par convolution (à l'exception notable des lois Gamma), le calcul de ces produits de convolution va passer par la discrétisation de la loi des montants de sinistres F . Le procédé de discrétisation va être fonction de ce pourquoi il est calculé de manière à être prudent. En effet le procédé sera différent selon que l'on calcule une probabilité de ruine ou une prime Stop-Loss de manière à ce que la quantité calculée dans le modèle collectif discrétisé soit toujours prudente par rapport à la quantité correspondante dans le modèle collectif théorique.

Notons Δ le pas de discrétisation et considérons les intervalles $I_i =]i\Delta, (i+1)\Delta]$. On notera f_i la masse de probabilité placée en $i\Delta$.

4.1. Discrétisation pour calculer une probabilité de ruine

Dans l'optique d'un calcul de probabilité de ruine ou de VaR, on va concentrer toute la masse de probabilité de l'intervalle I_i sur sa borne supérieure $(i+1)\Delta$, on a donc :

$$\begin{aligned}\Pr[X_\Delta = (i+1)\Delta] &= f_{i+1} \\ &= \Pr[X \in I_i] \\ &= F((i+1)\Delta) - F(i\Delta)\end{aligned}$$

On a ainsi $\forall t \geq 0, \Pr[X > x] \leq \Pr[X_\Delta > x] \Rightarrow G^{coll}(t) \geq G_\Delta^{coll}(t)$ ce qui assure que $S^{coll} \prec_{st} S_\Delta^{coll}$ (cf. notes de cours « Mesures et comparaison de risques »). On parle également de discrétisation en accord avec la VaR.

4.2. Discrétisation pour calculer une prime stop-loss

Dans l'optique d'un calcul de prime stop-loss, on va disperser la masse de probabilité de I_i à ses extrémités de manière à conserver la moyenne sur l'intervalle. Définissons pour $i \in \mathbf{N}$,

$$\begin{cases} f_i^+ = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \{(i+1)\Delta - x\} dF(x) = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \{F(x) - F(i\Delta)\} dx \\ f_{i+1}^- = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \{x - i\Delta\} dF(x) = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} \{F((i+1)\Delta) - F(x)\} dx. \end{cases}$$

On définit alors X_Δ par :

$$\begin{cases} \Pr[X_\Delta = 0] = f_0 = F(0) + f_0^+ \\ \Pr[X_\Delta = i\Delta] = f_i = f_i^- + f_i^+, i \geq 1. \end{cases}$$

Pour $k\Delta \leq x < (k+1)\Delta$, on a :

$$F^\Delta(x) = \Pr[X_\Delta \leq x] = \sum_{i=0}^k f_i = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} F(y) dy.$$

Ceci assure que :

$$\begin{cases} \mathbf{E}[S^{coll}] = \mathbf{E}[S_\Delta^{coll}] \\ \mathbf{E}[(S^{coll} - t)^+] \leq \mathbf{E}[(S_\Delta^{coll} - t)^+], \forall t \in \mathbf{R}. \end{cases}$$

Cette dernière inéquation assure que $S^{coll} \prec_{icx} S_\Delta^{coll}$ (cf. notes de cours « Mesures et comparaison de risques ») c'est à dire que la charge globale de sinistres dans le modèle collectif discrétisé est plus dangereuse que la charge globale dans le modèle collectif non discrétisé. On parle également de discrétisation en accord avec la TVaR.

5. Algorithme de Panjer

L'objet de ce paragraphe est de présenter la méthode la plus courante pour mettre en œuvre le modèle collectif.

5.1. Calcul direct de la fonction de répartition discrétisée

Supposons que les montants de sinistres $X_1, X_2, \dots$ sont à valeurs dans $\mathbf{N}^*$ et notons $f_i = \mathbf{Pr}[X = i]$, $f_0 = 0$. On a :

$$f_i^{*(k)} = \mathbf{Pr}\left[\sum_{j=1}^k X_j = i\right], i > 0, k \geq 0.$$

Comme les $X_1, X_2, \dots$ sont > 0 , les $f_i^{*(k)}$ peuvent être calculés par la récurrence suivante :

$$f_i^{*(k)} = \sum_{j=1}^{i-k+1} f_{i-j}^{*(k-1)} f_j, k > 0.$$

Comme S^{coll} est à valeurs dans $\mathbf{N}$ puisque c'est une somme (éventuellement nulle) d'entiers naturels strictement positifs, on notera dans la suite :

$$\begin{cases} p_k = \mathbf{Pr}[N = k] \\ g_i = \mathbf{Pr}[S^{coll} = i] \end{cases}$$

Alors

$$g_i = \sum_{k=0}^i p_k f_i^{*(k)},$$

et

$$G(x) = \sum_{j=0}^{[x]} g_j, x \geq 0,$$

où $[.]$ désigne l'opérateur partie entière.

Cette méthode nécessite un grand nombre d'opérations (calcul des produits de convolution par récurrence, etc.) et peut s'avérer inutilisable en pratique. Aussi au début des années 1980, Harry Panjer a proposé un algorithme récursif plus rapide pour calculer les g_j qui porte désormais son nom.

5.2. Algorithme de Panjer

L'algorithme de Panjer ne s'applique que si la loi de N fait partie de la classe de Panjer.

Définition 4 : La classe de Panjer contient toutes les lois de probabilité sur $\mathbf{N}$ qui satisfont

$$\text{la relation de récurrence : } p_k = \left(a + \frac{b}{k}\right) p_{k-1}, \quad k = 1, 2, 3, \dots$$

En particulier les lois de Poisson ($a = 0, b > 0$), binomiale négative ($0 < a < 1, a + b > 0$) et binomiale ($a < 0, b = -a(m + 1)$, pour un certain $m \in \mathbf{N}$) appartiennent à la classe de Panjer.

Proposition 2 : Si la loi de probabilité de N fait partie de la classe de Panjer et si $f_0 = 0$, les g_i satisfont la relation :

$$g_i = \begin{cases} p_0 & \text{si } i = 0 \\ \sum_{j=1}^i \left(a + b \frac{j}{i}\right) f_j g_{i-j} & \text{si } i \geq 1. \end{cases}$$

En particulier, si $N \sim P(\lambda)$ ($a = 0$ et $b = \lambda$), on a : $g_i = \begin{cases} \exp(-\lambda) & \text{si } i = 0 \\ \frac{\lambda}{i} \sum_{j=1}^i j f_j g_{i-j} & \text{si } i \geq 1 \end{cases}$.

Proposition 3 : Si la loi de probabilité de N fait partie de la classe de Panjer et si $f_0 > 0$, les g_i satisfont la relation :

$$g_i = \begin{cases} N(f_0) = \sum_{k=0}^{+\infty} p_k (f_0)^k & \text{si } i = 0 \\ \frac{1}{1 - a f_0} \sum_{j=1}^i \left(a + b \frac{j}{i}\right) f_j g_{i-j} & \text{si } i \geq 1. \end{cases}$$

Ce qui nous donne dans le cas de la loi de Poisson : $g_i = \begin{cases} \exp(-\lambda(f_0 - 1)) & \text{si } i = 0 \\ \frac{\lambda}{i} \sum_{j=1}^i j f_j g_{i-j} & \text{si } i \geq 1 \end{cases}$.

Le lecteur trouvera le détail des démonstrations de ces résultats dans DENUIT et CHARPENTIER [2004].

5.3. Calcul des primes stop-loss

L'algorithme de Panjer peut, en particulier, être utilisé pour calculer des primes stop-loss grâce au schéma suivant :

$$\Pr[S^{coll} > i] = \begin{cases} 1 & \text{si } i = -1 \\ \Pr[S^{coll} > i - 1] - g_i & \text{si } i \in \mathbf{N}, \end{cases}$$

et

$$\mathbf{E} \left[(S^{coll} - i)^+ \right] = \begin{cases} \mathbf{E} [S^{coll}] & \text{si } i = 0 \\ \mathbf{E} [(S^{coll} - i + 1)^+] - \mathbf{Pr} [S^{coll} > i - 1] & \text{si } i \in \mathbf{N}^*. \end{cases}$$

Par récurrence il vient :

$$\mathbf{Pr} [S^{coll} > i] = 1 - \sum_{j=0}^i g_j \text{ pour } i \in \mathbf{N},$$

et

$$\mathbf{E} [(S^{coll} - i)^+] = \mathbf{E} [S^{coll}] - \sum_{j=0}^{i-1} \mathbf{Pr} [S^{coll} > j].$$

Considérons à présent un niveau de franchise t (pas nécessairement un multiple du pas de discrétisation Δ), pour k entier naturel tel que $k\Delta \leq t < (k+1)\Delta$ on a :

$$\begin{aligned} \mathbf{E} [(S_{\Delta}^{coll} - t)^+] &= \int_{k\Delta}^{+\infty} \overline{G_{\Delta}^{coll}}(x) dx \\ &= \int_{k\Delta}^{+\infty} \overline{G_{\Delta}^{coll}}(x) dx - \int_{k\Delta}^t dx \\ &= \mathbf{E} [(S_{\Delta}^{coll} - k\Delta)^+] - (t - k\Delta) \overline{G_{\Delta}^{coll}}(k\Delta). \end{aligned}$$

L'erreur de discrétisation peut être bornée par :

$$0 \leq \mathbf{E} [(S_{\Delta}^{coll} - t)^+] - \mathbf{E} [(S^{coll} - t)^+] \leq \lambda \frac{\Delta}{4} \varepsilon_{\Delta},$$

où ε_{Δ} est la plus grande masse de probabilité accordée à un intervalle I_j , *i. e.*

$$\varepsilon_{\Delta} = \sup_{j \in \mathbf{N}} \{F((j+1)\Delta) - F(j\Delta)\}.$$

Au final, l'erreur totale résultat du passage du modèle individuel au modèle collectif et de la discrétisation dans le modèle collectif peut être bornée par :

$$\begin{aligned} 0 &\leq \mathbf{E} [(S_{\Delta}^{coll} - t)^+] - \mathbf{E} [(S^{ind} - t)^+] \\ &\leq \mathbf{E} [(S_{\Delta}^{coll} - t)^+] - \mathbf{E} [(S^{coll} - t)^+] + \mathbf{E} [(S^{coll} - t)^+] - \mathbf{E} [(S^{ind} - t)^+] \leq \lambda \frac{\Delta}{4} \varepsilon_{\Delta} + \frac{1}{2} \sum_{i=1}^n \beta_i q_i^2. \end{aligned}$$

6. Méthodes d'approximation de la loi de S^{coll}

Un certain nombre de techniques² permettent d'approximer la loi de la charge globale de sinistres dans le modèle collectif. Nous insistons ici sur deux méthodes particulières : la *Fast*

² Le lecteur trouvera un panorama de ces méthodes dans PARTRAT et BESSON [2004].

Fourier Transform (FFT) et l'approximation Gamma de Bowers. La première a le mérite d'être très répandue (notamment dans les logiciels statistiques) et la seconde permet de disposer d'une approximation dont les moments coïncident jusqu'à l'ordre 3.

6.1. Fast Fourier Transform

La technique FFT repose sur la transformée de Fourier inverse qui permet d'exprimer la loi d'une variable aléatoire arithmétique en fonction de sa fonction caractéristique.

Proposition 4 : *Si Y est une variable aléatoire arithmétique à valeurs dans $\{kh | k \in \mathbf{N}\}$ avec $h > 0$ fixé, en notant φ_Y sa fonction caractéristique, on a*

$$\Pr[Y = hk] = \frac{1}{2\pi} \int_0^{2\pi} e^{-ikt} \varphi_Y\left(\frac{t}{h}\right) dt \text{ pour } k \in \mathbf{N}.$$

Démonstration : Cf. PARTRAT et BESSON [2004]. $\square$

Comme en pratique, il est rarement possible de calculer l'intégrale dans l'expression précédente, la méthode FFT propose de découper l'intervalle $[0; 2\pi]$ en m sous-intervalles de même longueur par

$$p_k = \Pr[Y = hk] = \frac{1}{2\pi} \sum_{j=0}^{m-1} \int_{\frac{2\pi}{m}j}^{\frac{2\pi}{m}(j+1)} e^{-ikt} \varphi_Y\left(\frac{t}{h}\right) dt.$$

Puis en faisant l'approximation

$$\frac{1}{2\pi} \int_{\frac{2\pi}{m}j}^{\frac{2\pi}{m}(j+1)} e^{-ikt} \varphi_Y\left(\frac{t}{h}\right) dt \approx \frac{1}{m} e^{-\frac{2\pi}{m}ijk} \varphi_Y\left(\frac{2\pi j}{mh}\right),$$

et en posant

$$\begin{cases} \varphi_j = \varphi_Y\left(\frac{2\pi j}{mh}\right) & j = 0, 1, \dots, m-1 \\ w_{kj} = e^{-\frac{2\pi}{m}ijk} & k, j = 0, 1, \dots, m-1, \end{cases}$$

puis

$$\varphi^{(m)} = \begin{pmatrix} \varphi_0 \\ \vdots \\ \varphi_{m-1} \end{pmatrix}, \quad W = (w_{kj})_{k,j=0,1,\dots,m-1} \quad \text{et} \quad p^{(m)} = \begin{pmatrix} p_0 \\ \vdots \\ p_{m-1} \end{pmatrix},$$

l'approximation donne

$$p^{(m)} \approx \frac{1}{m} W \phi^{(m)} =_d \frac{1}{m} FFT^{-}(\phi).$$

La technique FFT repose sur une deuxième approximation qui découle de la proposition suivante.

Proposition 5 : Si $V = (v_{kj})_{j,k=0,\dots,m-1}$ où $v_{kj} = e^{\frac{2\pi}{m}ijk}$, on a $W^{-1} = \frac{1}{m} V$.

Cette proposition nous permet d'écrire

$$\phi^{(m)} \approx V p^{(m)} =_d FFT^{+}(p).$$

Au final la procédure d'obtention d'une loi approchée de S^{coll} par la méthode FFT contient quatre étapes :

- (1) Donnée des probabilités p_k de X après arithmétisation si nécessaire.
- (2) Calcul du vecteur $\phi^{(m)}$ par $FFT^{+}(p)$.
- (3) Calcul de $g_N(\phi)$ où g_N est la fonction génératrice des moments de la variable aléatoire « nombre de sinistres ».
- (4) Calcul de la loi approchée de S^{coll} par $p_{S^{coll}} \approx \frac{1}{m} FFT^{-}(g_N(\phi))$.

L'ensemble de cette procédure s'écrit donc

$$\frac{1}{m} FFT^{-}(g_N(FFT^{+}(p))).$$

6.2. Approximation Gamma de Bowers

L'approximation Gamma de Bowers consiste à approximer la charge totale de sinistres par une loi Gamma traduite. Cette modélisation permet de prendre en compte les trois premiers moments de S^{coll} : la charge totale de sinistres va être approchée par une loi Gamma traduite dont les trois premiers moments coïncident avec ceux de S^{coll} .

Propriété 1 : (Rappels sur la loi Gamma) Si $X \sim \gamma(v, \beta)$, on a :

$$\begin{aligned} f(x) &= \frac{\beta^v}{\Gamma(v)} x^{v-1} e^{-\beta x}, \\ F(x) &= \Gamma(v, \beta x), \\ m_k &= \frac{v(v+1)\dots(v+k-1)}{\beta^k}, \end{aligned}$$

où $\Gamma(v, \cdot)$ est la fonction Γ -incomplète qui à tout couple (v, x) dans $\mathbf{R}_+^2$ associe $\frac{1}{\Gamma(v)} \int_0^x e^{-t} t^{v-1} dt$ où $\Gamma(v) = \int_0^{+\infty} e^{-t} t^{v-1} dt$.

Le lecteur trouvera dans PARTRAT et BESSON [2004] les principales propriétés des fonctions Γ et Γ -incomplète.

Définition 5 : Une variable aléatoire Z est distribuée selon une loi Gamma traduite de paramètre (v, β, x_0) si $Z - x_0$ suit une loi Gamma $\gamma(v, \beta)$.

La loi Gamma traduite est définie par trois paramètres qui vont être choisis de manière à ce que l'approximation de S^{coll} ait les mêmes trois premiers moments que S^{coll} , i. e.

$$\mathbf{E}[S] = x_0 + \frac{v}{\beta}, \quad \mathbf{Var}[S] = \frac{v}{\beta^2}, \quad \mu_3(S) = \frac{2v}{\beta^3}.$$

Ce qui revient à choisir les paramètres

$$\beta = \frac{2 \mathbf{Var}[S]}{\mu_3(S)}, \quad v = \frac{4(\mathbf{Var}[S])^3}{\mu_3^2(S)}, \quad x_0 = \mathbf{E}[S] - 2 \frac{(\mathbf{Var}[S])^2}{\mu_3(S)}.$$

Ces paramètres pourront être estimés à partir des estimateurs empiriques des trois premiers moments de S^{coll} .

L'approximation Gamma consiste à approcher la fonction de répartition F_S de S^{coll} par :

$$F_S(x) \approx \Gamma(v, \beta(x - x_0)) \text{ pour } x > x_0,$$

En se fondant sur cette approximation Gamma, Bowers a proposé d'approximer la fonction de répartition de S^{coll} par :

$$F_S(x) \approx F_S^{Bowers}(x) = G\left(\frac{\mu_S}{\sigma_S^2} x\right) \text{ si } \mu_3(\beta S) = \left(\frac{\mu_S}{\sigma_S^2}\right)^3 \mu_3(S),$$

où :

- $G(x) = \Gamma(v, x) - \frac{1}{6} [\mu_3(\beta S) - 2v] e^{-x} x^v \left[\frac{x^2}{\Gamma(v+3)} - \frac{2x}{\Gamma(v+2)} + \frac{1}{\Gamma(v+1)} \right],$
- $\beta = \frac{\mu_S}{\sigma_S^2},$
- $\alpha = \frac{\mu_S^2}{\sigma_S^2}.$

Ce résultat est démontré à l'aide de propriétés sur les polynômes orthogonaux de Laguerre dans PARTRAT et BESSON [2004].

Bibliographie

DENUIT M. [2003] *Modèle collectif de théorie du risque*. Notes de cours, ISFA.

DENUIT M., CHARPENTIER A. [2004] *Mathématiques de l'assurance non-vie. Tome 1 : principes fondamentaux de théorie du risque*. Economica.

GOURIEROUX CH. [1999] *Statistique de l'assurance*. Economica

PARTRAT CH., BESSON J.L. [2004] *Assurance non-vie. Modélisation, simulation*. Economica.

Exercices

Exercice 1 : Considérons un portefeuille constitué de trois risques générant des charges annuelles de sinistres de montant S_1, S_2 et S_3 , dont les lois de probabilité sont données dans le tableau suivant.

Valeurs de $\Pr[S_j = k]$			
j	$k = 0$	$k = 1$	$k = 2$
1	0,4	0,3	0,3
2	0,6	0,1	0,3
3	0,4	0,4	0,2

Décrivez l'approximation collective S^{coll} de $S^{ind} = S_1 + S_2 + S_3$ en vue de calculer des primes Stop-Loss $\mathbf{E}[(S^{ind} - t)^+]$.

Exercice 2 : Considérons le modèle collectif $S^{coll} = \sum_{i=1}^N X_i$ où N est une v.a. de loi de Poisson de paramètre λ et où les montants $X_1, X_2, \dots$ sont indépendants et distribués selon une loi exponentielle de paramètre θ . En vue de calculer la prime Stop-Loss $\mathbf{E}[(S^{coll} - t)^+]$ à l'aide de l'algorithme de Panjer, il est nécessaire de discrétiser les X_i . Déterminez le pas de discrétisation Δ afin de limiter l'erreur correspondante à 10 % de la prime pure $\mathbf{E}[S]$.